

Can Europe escape the doldrums?

The difficulty and delay in ratifying the Maastricht Treaty cast doubts on the soundness of the project and demand a greater sense of vision from those who would make Europe a reality.

THIS was to have been Europe's year. Not only would the single market be in effect, providing the economic benefits offered almost immediately by the Treaty of Rome, but most of the twelve member states of the European Communities (EC) would have ratified the Maastricht Treaty, mapping out progress towards "ever closer union" (as the treaty puts it). Although the single market is in being, compliance with its terms cannot yet be verified. The more substantial worry now is not so much that two countries (Britain and Denmark) have not yet ratified the treaty (and may not do so), but the suspicion that many of the governments ratifying the treaty last year would not do so this. In Europe falling apart?

The signs are not encouraging. From one end of Europe to the other, governments are paralysed by hesitancy. Germany is paying for reunification the hard way, beset by more immigrants than it can accommodate and heading for an election as well. Italy is stymied by corruption in public life. France, with a new government, has yet to declare its allegiance to EC institutions such as the Common Agricultural Policy and will dither for some time yet over the tentative deal between the EC and the United States over the proposed new arrangement under the General Agreement on Tariffs and Trade (GATT). Britain's diplomatic and political energy is dissipated by the year-long struggle to get the Maastricht bill through the House of Commons. The tangible proof of Europe's disunity is the lamentable indecision of the North Atlantic Treaty Organization over its role in the remnants of Yugoslavia.

None of this is a surprise. When it became plain that Maastricht would stick in a near-majority of European throats, this journal said it should be renegotiated. That has been borne out by much that has happened since. Thus the British government's defence of Maastricht in the House of Commons (to be followed by a similar battle in the House of Lords), on which it stakes its political survival, runs roughly on the lines: "Sure, it's a poor treaty, but ratification is symbolic only". Literally, that is true: in the monetary field, for example, there is now little chance that there will be a European Central Bank between four and six years from now, as Maastricht specifies. (Paradoxically, the goal would be more easily achievable if the deadline were tomorrow or the day afterwards, which could happen if the member states regarded the obdurate recession as a crisis best dealt with by monetary union, but there is little chance of that either.) But why should Europe be brought to a standstill because politicians have decided to battle over a piece of paper when

they have more urgent matters to attend to?

Governments have painted themselves into this corner in the simple but mistaken belief that it is possible to build a coherent federation (such as the United States) or confederation (such as Switzerland) by legislating for free trade between member states and then giving a central executive committee (the European Commission) powers to ensure that traders compete on equal terms. By abreaction, most of the twelve governments have come to regard Brussels as a kind of jousting ring at which their representatives battle for national prizes, not for the broader European interest. Even Franco-German solidarity seems to be wearing thin. So who should be surprised that so little has been heard in the past few years of extending the remit of the EC towards the east? So long as people believe that the European venture is a zero-sum game, they will be fearful that new members' gains will be their loss. But was it not once believed, as recently as 1955, that the whole would be greater than the sum of the parts?

The present crisis in European affairs could damage the prospects for science in Europe, which have been marvellously stimulated by the past few decades of growing collaboration. The commission, to its credit, has consistently made the right noises in that and some other directions. It has, for example, been a generous source of travel funds, while not all of its spending on the current chaotic fellowship programme will in the end be wasted. With further iteration, Europe's programmes in support of basic research could be distinctively enlivening for European science and thus for Europe: the whole would indeed be larger than the sum of its parts. But that will not happen if Europe tears itself apart. Science, of course, will not be the only or even the most important casualty of such an outcome. It is merely an illustration of why European governments should try to rise above the narrow nationalism that now grips them. □

Venter's venture

Much of the work on the human genome may be done sooner than expected, but there will still be much to do.

DR Craig Venter's claim, at the *Nature Genetics* conference two weeks ago, that he may have listed 95 per cent of the genes in the human genome within the next eighteen months will, if proved true, turn many Human Genome Projects on



their heads. Hitherto, people have been taking the long view, planning to practice for a few years on nematodes, *Arabidopsis* and *Drosophila*, before turning their attention, and the techniques learned in the earlier stages, to the human genome. On that leisurely timetable, a nucleotide sequence of the human genome would be complete sometime in the first decade of the next century. Does Venter's gallop through the genome render the slower programmes redundant?

The simple truth, of course, is that it does not. Venter isolates only tags, represented by nucleotide sequences of variable length, from the ends of genes. Indeed, it is worse than that: the tags are DNA molecules complementary to the ends of RNA molecules recovered from the cytoplasm of a cell. While they may represent the genes active in particular tissues, they can say nothing about the parts of those genes which are not expressed (called introns) nor about the apparently meaningless DNA in which many genes are embedded (called 'junk', Crick's word). Even so, Venter's approach offers quick way of labelling active genes (whence the name 'tag') which can be located within the genome and then fished out from it so that their complete sequence can, if necessary, be determined. That is grist for the mills of those who want to understand the particular functions of particular genes, which is how medical practice will be eventually improved. So the more complicated truth about the complementary DNA (cDNA) technique is that it is a threat to the classical human genome projects; agencies will more readily fund the former than the latter.

That would be mistaken. Several important features of the human and any other genome might then indefinitely be mysteries. That there should be introns within genes has been a puzzle since they were first recognized in 1980. Are they functional, and if so how, or perhaps vestigial and, if so, with what meaning? Is the 'junk' really junk, and even so, may there be evolutionary lesson to be learned from the comparison of one creature's junk with another's? Then, since the human genome (or any other) contains not just a specification of the components of each cell but a recipe for making them, must not the detailed arrangement of the genome embody not merely the structure of all of an organism's molecules but the work-plan for generating them as and when required? And who would say that answers to such questions as these would be without medical benefit?

What this implies is that the two approaches to the structure of the human genome are neither in conflict nor complementary but different. The classical human genome project, figuratively (only) that of starting from one end and working through to the other 3 billion base-pairs later, will yield information that the cDNA technique cannot. To abandon the first because the second will yield immediately useful information would be like trying to teach a person a foreign language by instructing him only on the meanings of the nouns. The difficulty, at least until the dust has settled, is that the cDNA sequences, which would be of great value to the larger project, are likely to be generated most rapidly under the umbrellas of companies and may not be published until patent rights are granted, whenever that may be. That would be a great waste of effort. □

Vaccine compromise

The money about to be wasted on a clinical trial for AIDS vaccines could be better spent on basic science.

IN what at first looked like a triumph of military might over common sense and sound scientific judgement, the US Army decided last week to go ahead with a \$20 million trial of a "therapeutic" AIDS vaccine despite objections from the US National Institutes of Health (NIH) and the Food and Drug Administration (FDA) (see page 581). The vaccine, based on the gp160 subunit of the human immunodeficiency virus (HIV), is promoted by its manufacturer (a small Connecticut company called MicroGeneSys) as a significant new therapy for AIDS. The trial was mandated by Congress last year only after extensive lobbying by a company that has aggressively promoted its product.

But even as word of the Army's decision to proceed despite scientific opposition became known, NIH and FDA officials, as well as AIDS activists, were protesting to the White House and the Department of Defense that any trial should include several candidate therapeutic vaccines now being studied, not just the one from MicroGeneSys. Perhaps surprisingly, the NIH and FDA seem to have carried the day and their apparent success in demanding a multi-vaccine trial is seen as the virtuous triumph of science.

In a way it is. If the United States is going to spend \$20 million or more to test these vaccines on thousands of patients, it is better to evaluate all candidate vaccines. The NIH feels so strongly about a multi-vaccine approach that would include patients from minority groups, intravenous drug users and others with AIDS who are not in the military that it has volunteered to add money to the pot, thereby taking the wind out of the Army's argument that a broad trial would cost more than the \$20 million Congress has allocated.

But the real issue is whether any therapeutic vaccine should proceed. Taking the MicroGeneSys case as an example, the argument is that the gp160 subunit might lead to an immune stimulation sufficient to produce a clinically meaningful halt to HIV in people who are already infected. In this sense, the therapeutic vaccine is more like a drug than a traditional vaccine, which is meant to prevent infection in the first place. But the accumulation of new data, many of them dependent on refined techniques, such as PCR (polymerase chain reaction), for finding HIV in people who are infected but not yet afflicted with overt AIDS, suggests that, within weeks of infection, HIV in massive quantities has seeded the lymph nodes and is present in blood (see *Nature*, 362, 355; 1993). A therapeutic vaccine may simply be too little too late. Certainly trials enrolling thousands of patients are not yet in order.

At this point, NIH, FDA, the Army and Congress should do the right thing and reevaluate the entire premise of this \$20 million trial which, by implication, holds the promise of a significant advance in the treatment of AIDS. If they do not, they stand accused of playing politics with people suffering from a complex and still lethal disease. □

Clinton prepares to sign biodiversity treaty to celebrate Earth Day

Washington. US President Bill Clinton is expected in the next few days to announce his intention to sign the international agreement on biodiversity protection that former president George Bush rejected at last June's Earth summit in Rio de Janeiro. The announcement may come as part of a package of environmental initiatives to be unveiled on 22 April to celebrate Earth Day.

The Bush administration said that it was opposed to the treaty on the basis of ambiguous language about ownership of patent rights for products derived from indigenous plants and animals. It is not clear whether companies would be forced to share the rights with or even transfer them to the countries that owned the biological resources. Bush said that such a policy would be harmful to US biotechnology and pharmaceutical companies.

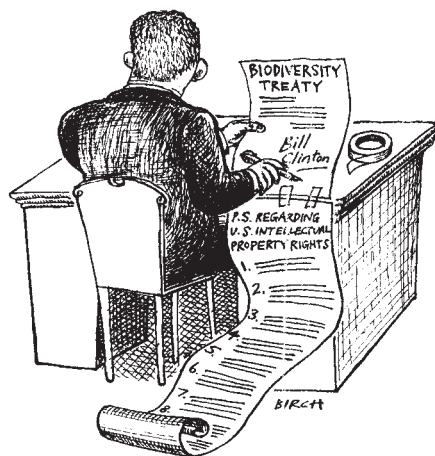
Several companies that sided with the Bush administration at the time have since withdrawn their opposition to the treaty, saying that they were pressured into supporting the decision without having studied the issue thoroughly. Bush's concern about intellectual property rights is believed to have been a smokescreen: a more important reason was the one given by the since-disbanded White House Council on Competitiveness, which was concerned that increased protection for endangered species would make it harder for US companies to operate at home and abroad.

However, even those who favour the treaty admit that serious problems remain over what Richard Mott of the World Wildlife Fund calls the "inartfully and ambiguously drafted" convention. In one paragraph, the treaty seems to provide for "the adequate and effective protection of intellectual property rights", while elsewhere it implies that signatory countries — and private companies within those countries — would have to share whatever technology is derived from native biological resources.

Another point of controversy is the implication within the convention that wealthy developed nations would be required to make fixed payments, rather than voluntary contributions, to developing countries for biodiversity protection programmes. A related matter is how those funds would be distributed.

In recent months, a coalition including Genentech and Merck and Co., along with the World Wildlife Fund and the World Resources Institute, have been discussing with the White House ways for the United States to sign the convention without sacrificing intellectual property rights. Their solution is for the United States to attach a

rider clarifying its interpretation of any controversial provisions. "Vagueness can be your worst enemy or your best friend", says Ray Briscuso of the Industrial Biotechnology Association, which originally opposed



the treaty but now supports attaching a clarification to the president's signature.

The European Communities (EC) also signed the biodiversity convention at Rio but remain concerned about the intellectual property rights issue, and a similar rider spelling out its interpretation is expected to accompany ratification. Fourteen countries have already ratified the treaty, almost half of the 30 needed to bring it into force.

EC environmental officials last week met

Vice President Al Gore and his staff in Washington, where they discussed the biodiversity convention as well as European plans for a carbon tax and Clinton's proposed energy tax as methods of reducing worldwide emissions of greenhouse gases. Tougher worldwide restrictions on greenhouse gases are among the international environmental matters under review by the White House National Security Council. The review, which began in March, includes wetlands, oceans, trade, population and sustainable development.

Of all the international issues on the table, signing the biodiversity accord was seen as one of the easiest to solve. "I think Clinton wants to do something notable [for Earth Day], and that's a very likely candidate", says Mott.

Even if the United States signs the convention before the one-year deadline in early June, it is not clear what it must do to conserve its biological resources. The treaty is vague about the need for signatories to protect their own biodiversity.

A presidential announcement would reflect the convention's value as a symbolic step towards preserving global biodiversity. What will really make a difference are protocols established after the biodiversity convention comes into force that are similar to the 1987 Montreal Protocol restricting the use of chlorofluorocarbons to protect ozone. "Anything with teeth is going to be in the protocols", says Barbara Bramble of the National Wildlife Federation.

Tony Reichhardt

Synergen trims sails to survive

Washington. The US biotechnology company Synergen Inc., reeling from disappointing clinical results that have driven down the price of its stock by more than two-thirds, last week replaced its chief executive and reorganized management.

Board chairman and company co-founder Larry Soll has replaced Jon Saxe as chief executive officer and four executive vice presidents will share the duties of president, a post also held by Saxe, as the company, based in Boulder, Colorado, tries to recover from recent poor results of trials of its anti-sepsis drug Anril. Michael Catalano, vice president of clinical research, will become a technical consultant to the company. Synergen plans a second phase-III trial with Anril, an interleukin-1 antagonist, that will focus on "seriously ill" patients in the hope of showing more dramatic gains.

One silver lining for Synergen is its \$220

million in capital, which will allow it to conduct a follow-up trial with Anril as well as to pursue plans to market the drug in Europe and to begin joint trials with Syntex on a drug to treat amyotrophic lateral sclerosis, or Lou Gehrig's disease. A spokesman says that the company may save money by terminating some research, for example work on an elastase inhibitor for cystic fibrosis.

Synergen believes that data showing a reduction of 22 per cent in mortality for the subset of most critically ill sepsis patients is sufficient basis for a licence in Europe, although mortality for the entire patient population tested was not significantly different from placebo control rates. With no other products likely to be ready before the late 1990s, Synergen's best bet may be to use its capital to buy a smaller company needing money to develop a product that is closer to market.

Jenna Roberts

Spain expands despite financial and logistical hurdles

Madrid. Spain is hoping that money from industry and the European Communities will help it to expand research facilities in the provinces despite a shrinking national science budget and formidable administrative hurdles. Among the expected beneficiaries of that support are two institutes for plant biology and for food technology being built in Valencia and a cell and molecular biology institute that will soon open in Seville. But the fate of the newly inaugurated National Centre for Biotechnology (CNB) in Madrid — whose opening was delayed for several years and which is still not fully functional — casts doubt on prospects for a smooth launch of the new centres of excellence.

Spanish science has made significant progress in the past 15 years, with an annual growth in budget of 18 per cent, to Pts600 billion (US\$5 billion). The early days of expansion spawned the concept of building world-class centres of excellence, including the National Centre for Microelectronics in Barcelona, which opened in 1990, and the CNB, which began operations last September.

This growth has occurred despite the continuing concentration of scientific funds and personnel in Madrid and Barcelona. For example, in 1988 only a third of all research funds from the Ministry of Health went to the rest of the country. The Spanish government, determined to change this pattern, decided with Spain's research council (CSIC) to build institutes outside the two major cities with money from other sources. Both institutes in Valencia were built with funds obtained from the European Commission (EC) on land donated by the local university. The Seville Institute, now called the Andalusian Laboratory of Biology, hopes to get money from the EC as well as from the regional government of Andalusia.

The new institutes hope to avoid the problems that have beset the biotechnology institute, which has suffered from the ingrained inefficiencies resulting in part from a shift from the highly centralized adminis-

tration in the early 1980s to a more complicated bureaucracy that provides local and regional administrations as well as universities with their own budgets. The CNB must deal with all three parties.

CNB's first director, the Spanish-speaking, British-born biologist Michael Parkhouse, resigned in 1990 after four years because of the delays and the related uncertainties of his one-year contracts. The same problems precipitated the resignation after two years of the second director, José López Carrascosa. Both men complained of the time spent on arranging for continued funding for the centre.

The appointment of scientific staff is another major problem for the CNB. Although the system was created to ensure a degree of autonomy from CSIC, tenured laboratory chiefs (of whom there will be 36) must be approved first by a scientific advisory board, then by a national board with representatives of government and industry, and finally by CSIC itself. Once appointed, laboratory chiefs face the uncertainty of having their scientific plans altered by layers of bureaucracy. (The selection of ten of these positions has been postponed until new scientific advisory boards have been established.)

The current director, Mariano Esteban, a molecular biologist appointed last August, is well aware of these problems, with two department chiefs — each responsible for eight laboratories — having resigned in the past few weeks in disputes over scientific priorities and staff appointments. Francisco Malpartida, who was chief of CNB's department of microbiology, says that "some primary research lines suddenly became very low priority and so scientists who had previously been appointed could not join the centre".

Discontent also extends further down the ranks. Marisa Toribio, a CSIC *investigador* — the first step towards tenure — was offered a post as laboratory chief but has decided to remain in an institute near the CNB. She says the move would have restricted her scientific independence and left her with an uncertain future.

Esteban must also grapple with the effects of the recession, which has led to cuts in the Spanish science budget and a freeze on all civil service appointments for 1993. The centre's operating budget, originally expected to reach Pts500 million, has been reduced and runs out in summer, and industrial support is being sought. The ban on new appointments has cre-

ated a situation in which the centre, the most modern and sophisticated in Spain with facilities for 350 scientists, stands virtually idle for want of the necessary technical support.

Officials at the three new institutes say they have learned from CNB's troubles. "We will try to avoid a complicated decision-making processes at all levels," says Enrique Cerdá, a university professor in Seville who is helping to set up the cell and molecular biology institute. "Researchers will be recruited only on scientific merit and appointments will be reviewed on a five-year basis [instead of those concerned being offered tenure]."

Cerdá has a good chance of success because his institute is at present independent of CSIC, but the Valencia institutes must abide by CSIC rules. Last year, for example, CSIC offered only 78 tenured positions for its hundred institutes, making it unlikely that the Valencia institutes will get "the [necessary] input of fresh scientific persons and ideas", says Daniel Ramón, an *investigador* who is moving to one of the new institutes.

Despite these problems, Spanish scientists still expect to have as much autonomy as their counterparts in other European countries. And they remain convinced that the new facilities will in the long run raise the quality of Spanish science.

Oscar G. Segurado

Station redesign clouds NASA budget for space science

Washington. The amount of funding available next year for several proposed new space science and technology projects — from low-cost planetary missions to small Earth probes and lunar orbiting spacecraft — awaits a decision by the US National Aeronautics and Space Administration (NASA) and its international partners on the redesign of space station Freedom (see *Nature* 362, 484; 1993).

President Bill Clinton's budget released last week includes a request for \$15.625 billion for NASA in fiscal year 1994, with a total of \$2.3 billion for the station and for new science and technology initiatives. Last autumn, Congress appropriated \$2.2 billion in 1993 for the station alone. The White House has ordered NASA to study three options for building the station — with costs of \$5 billion, \$7 billion or \$9 billion over five years — and to consider building it jointly with Russia. Until the redesign is complete, however, the space agency will not know how much is left for new projects.

Tony Reichhardt



Biotechnology centre struggles in Madrid.

Russian nuclear accident opens up military complex

Moscow. The chemical explosion last week in a reprocessing tank in the Russian city of Tomsk-7 has also blown the cover of a formerly secret city that until recently did not officially exist.

The plutonium reprocessing plant is at the city of Tomsk-7, built in the 1940s to produce plutonium for Soviet nuclear weapons. It now emerges that three of the six reactors are still working. Soviet authorities had said the purpose of the plant was to provide fuel for civilian nuclear power plants. The city is situated on the Tom River bank about 20 km north of the Siberian town of Tomsk, which is 3,000 km east of Moscow.

The explosion occurred on 6 April in a stainless steel holding tank within minutes of the addition of nitric acid as a solvent. The explosion blew a reinforced concrete lid through the roof of the plant.

There is considerable debate about the contents of the tank and the potential threat to residents and to the environment from the radiation released in the explosion. The ministries in charge say that the tank contained uranium-235 and plutonium-237 but not plutonium-239, which is more toxic. That, however, implies that the nuclear fuel in the reactors contained no uranium-238, the chief constituent of natural uranium.

Alexei Yablokov, ecology and public health adviser to President Boris Yeltsin, says that he will ask the government "to conduct a thorough check of all potential radiation hazards in Russia, including military ones". Yablokov believes that Russian officials have underestimated the doses to which firefighters were exposed and that levels inside the building were many times higher than what was reported outside. One rescue worker is reported to be under treatment.

The initial official report, denied the following day, said that the solution con-

tained plutonium and uranium salts and that some amount of plutonium-237 and uranium-235 were released into the atmosphere. There were similarly conflicting reports about the existence of a cloud said to be moving toward a nearby town.

Viktor Mikhailov, minister for nuclear energy, said that plutonium has not been produced at Tomsk-7 in the past year. A subordinate contradicted his statement but said that no plutonium was released in the explosion. Lidiya Popova, an expert on nuclear physics at the Socio-Ecological Union, an environmental organization, who twice visited Tomsk-7 last year, says that plutonium-239 is common at military complexes such as Tomsk-7.

The accident occurred just before the seventh anniversary of the disaster at the Chernobyl nuclear power plant, but this 'nuclear spring' was announced within a few hours. It is also the first time that such an

event at a military nuclear plant was made public when it happened instead of months or years later.

Yelena Bonner, widow of Andrei Sakharov, says that the incident shows that "we need Western experts to assess the situation. The fact that this plant was producing secret weapons is the most worrying. International controls need to be increased." Others have said that the two accidents demonstrate that every part of the chain in handling nuclear material is important and that the state must tighten its control over the process.

Nevertheless, the average Russian citizen seems less concerned about the potential threat to public health posed by nuclear weapons than about the current economic crisis — in particular, last week the price of petrol was doubled, from 40 to 80 rubles, an increase that Yeltsin has promised to overrule.

Monthly Nature, Moscow

Japan increases share of Human Frontier awards

Tokyo. Japan's scientists performed well in the latest round of awards from the Strasbourg-based Human Frontier Science Program (HFSP), winning more fellowships than scientists of any other country and trailing only the United States in the number of grant recipients. But the awards for international research on the brain and biological functions announced last week also show that Japan is the least popular destination for bright young scientists.

The Frontier programme, which this year has an annual budget of about US\$40 million provided largely by Japan, was established in 1989 on Japan's initiative with the backing of the seven Western summit nations (Japan, United States, Canada, United Kingdom, France, Germany and Italy) and the European Communities (EC). Applications for grants and fellowships, which are open to scientists around the world, are screened by international peer review.

Competition is severe: the 191 successful grant recipients represent only 12–13 per cent of the applications, which require proposals from international teams of scientists. And more than 500 young scientists applied for the 152 long-term fellowships awarded this year.

The latest round (see table) shows that the United States once again has the largest number of grant participants, followed by Japan and Britain. In addition, six of the 42 principal scientists in charge of the international teams are Japanese, as are 35 of the 152 long-term postdoctoral fellows. But only one of the 152 fellows has chosen to work in Japan, with a majority choosing the United States and most of the rest headed for Europe.

Japanese government officials and scientists are embarrassed by the low level of interest in Japan, citing problems with language and housing as well as deteriorating facilities at government laboratories and universities. Applicants may also be deterred by restricted career opportunities for foreign scientists and a lack of information about Japan's research organizations.

David Swinbanks

Debris no hazard to space flights

Munich. Space debris does not constitute a hazard to space missions, and it is neither technically practical nor economically feasible to remove the 7,000 objects in space left by earlier missions, according to 250 scientists from 17 countries who met last week in Darmstadt, Germany.

The participants agreed that prevention was the best approach and suggested ways to keep debris levels within safe limits, including boosting satellites into higher orbits at the end of their mission and reducing the amount of material that accompanies the intended payload into orbit. The group also agreed to continue monitoring the situation and to meet again in a few years. **A.A.**

HFSP scorecard

Grants

Country	Researchers	Leaders
Canada	11	4
France	15	6
Germany	19	4
Italy	7	2
Japan	34	6
Switzerland	10	3
United Kingdom	23	4
United States	52	12
Rest of EC	4	1
Others	16	0

Fellowships

Country	No. of fellows	No. hosted
Canada	6	5
France	10	9
Germany	15	10
Italy	4	1
Japan	35	1
Switzerland	5	5
United Kingdom	21	19
United States	15	95
Rest of EC	17	4
Others	24	3

British universities object to idea of year-round terms

Birmingham. British universities have reacted warily to the suggestion of changes in the academic year that would include year-round teaching. The government has endorsed the proposal as a way to save money, but many universities fear it would be damaging to research.

The implications of restructuring the traditional three-term academic year, which starts in late September and runs until the end of June, are being analysed by a committee chaired by Lord Flowers, a former rector of Imperial College in London and vice-chancellor of the University of London. The committee's investigation, whose preliminary report was published last week, has been prompted by tests of new academic schedules at several universities, including two 15-week terms instead of three of 10 weeks.

Addressing a conference in Birmingham organized by the Higher Education Funding Council for England (HEFCE) — the new

body responsible since 1 April for distributing government grants to England's 140 universities and other higher education institutions — Flowers said that spreading the teaching load more evenly across the year could help to alleviate overcrowding. In addition, if extra teaching resources were made available, the number of students could be increased by introducing an extra term or semester over the summer.

The government is keen for universities to investigate such new teaching patterns in keeping with its commitment to increase participation in higher education from 28 per cent to 33 per cent of school leavers by the end of the decade. John Patten, the Secretary of State for Education, says that there was "no question" of a single blueprint being imposed on all universities but that he is "optimistic that there is scope for considerable progress".

But senior officials from some universities say that their combined capacity for

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Patten expects progress.

teaching and research is already close to its limits. "The whole concept is ludicrous, certainly for research-intensive universities", says Derek Roberts, provost of University College, London.

Officials at both the Universities of Oxford and Cambridge say that they are unlikely to implement the type of radical changes envisaged in the Flowers report. Equally strong opposition — including the threat of strike action — has also come from the Association of University Teachers, which argues that, in the absence of substantial extra funds, requiring teaching during the summer months would undermine the opportunity to carry out research.

Responding to such criticisms, Flowers emphasizes that the changes outlined in his report would not necessarily be harmful to research. For example, he suggests that some institutions would benefit from distributing research more evenly throughout the academic year rather than concentrating it in the summer vacation.

Flowers also argued forcefully against the introduction of two-year degree courses. Although these might be popular with certain students (for example, mature students keen to reduce the time spent studying for a degree) he expressed doubts that a two-year degree "would provide sufficient time for reflection, field studies and laboratory work" for the majority.

In his address to the funding council, Patten said that he was strongly opposed for both intellectual and financial reasons to the widespread expansion of four-year undergraduate courses, even though these could be justified in exceptional circumstances. He deplores the trend towards adding a fourth year to courses in science subjects to compensate for inadequate teaching in schools, saying that the solution is better science teaching rather than lengthening the undergraduate course. "Longer higher education courses are a luxury we simply cannot afford" he said, adding that the funding councils have been asked to devise measures to discourage any increase in the average length of courses. **David Dickson**

Changes urged in reimbursement

London. The vice-chancellors of Britain's principal research universities have asked William Waldegrave, the cabinet minister responsible for science, to make urgent changes in the procedures to claim back funds transferred last year from the former Universities Funding Council (UFC) to the research councils. The government appears eager to resolve the problem, which involves the implementation of new rules on reimbursement for research-related expenses.

The request follows evidence produced by a working group of university finance directors that as much as £25 million (US\$37 million) of the £87 million being transferred during the current financial year may not have been allocated as intended. In particular, much of the money intended to cover the 'new' costs of technical staff and other exceptional items appears to have gone for 'old' costs such as research assistantships and major pieces of research equipment.

Two years ago, the government decided that research councils should pay the full costs of the research they sponsor at universities rather than relying on universities to cover some of the costs through their UFC grants. The government had intended that the money transferred to the research councils should make its way back to universities.

Detailed analysis of the 586 awards received by a sample of 15 universities from the five research councils since the beginning of the year provided no evidence that the transferred money has not been returned.

However the group did find an apparent shortfall of almost £3 million in the staff costs which the universities could have expected to receive back from the research councils under the new arrangements. In addition, only £1.3 million (rather than an anticipated £8.3 million) had been requested for the new types of non-staff costs that can now be claimed.

One reason for the discrepancies is that many academics have not been making claims under their grant applications for all the costs which would previously have been covered from university funds. University scientists, in turn, have complained that money can be claimed for technical assistants only if this is for more than 20 per cent of their time, and that research councils have apparently been disallowing a greater proportion of the 'new costs' than of the 'old costs'.

The two sides are planning to meet soon under the aegis of the Advisory Board for the Research Councils to iron out their differences. At the same time, immediate cash flow problems facing several universities as a result of the discrepancies has led to demands for action as soon as possible. In particular, universities, while accepting the need to ensure that their staff are better informed about the new procedures, want the research councils to eliminate the threshold on the amount they can claim for staff salaries and to reduce the thresholds applied to non-staff costs.

David Dickson

US test of AIDS vaccines is broadened beyond VaxSyn

Washington. A controversial plan to conduct large-scale clinical trials of a single therapeutic AIDS vaccine has been modified to compare the efficacy of three vaccines and will be conducted by the National Institutes of Health (NIH) instead of the US Army.

The US Departments of Defense and of Health and Human Services, NIH's parent agency, agreed last week to spend \$23 million to test the efficacy of three potential vaccines on 9,000 people after the Army had made known its intention to proceed with a single trial of VaxSyn, a gp160 therapeutic vaccine produced by the Connecticut biotechnology company MicroGeneSys. Last autumn, the company successfully lobbied Congress to appropriate \$20 million for that purpose. NIH has agreed to provide the additional money for the expanded trials.

The new plan is a victory for scientists and AIDS activists, who argued strongly that it would be premature to conduct a large-scale trial of VaxSyn. A panel co-chaired by NIH director Bernadine Healy and David Kessler, head of the Food and Drug Administration, last winter recommended the testing of several potential vaccines, but the Army was not required to heed their advice.

An announcement on the agreement, due late last week, was delayed as the Clinton administration tried to decide how to recon-

cile language in last year's defence appropriations bill with its decision that the NIH should spend the money instead of the Army. "The White House is still trying to figure this one out", says a Senate staff member. "It is trying to steer a middle course between what the law says and what the NIH and the DOD want to do."

The appropriations subcommittee is expected to discuss the change when the Senate returns next week from its two-week Easter recess. One possible solution is for the Army to award a contract to NIH to conduct the trials.

AIDS activists are confident that multi-vaccine trials will soon begin. "It's what we wanted", says Gregg Gonsalves of the Treatment Action Group in New York.

But MicroGeneSys president Frank Volvovitz expresses scepticism that a multiple trial large enough to provide conclusive results can be done for \$23 million. Volvovitz was scheduled to meet Department of Defense officials earlier this week to press his case for the single vaccine trial.

The \$23-million trial is thought to involve giving each vaccine to 3,000 people, against the Army's original plan to test VaxSyn on 10,000 people. Either approach would represent the largest clinical trials so far of the ability of therapeutic vaccines to retard the onset of full-blown AIDS on patients who are HIV-positive.

Colin Macilwain

NIH genome centre begins build-up

Washington. Twenty senior genome researchers are expected to join Francis Collins, the new director of the National Center for Human Genome Research at the US National Institutes of Health (NIH), in building an intramural programme that will focus on applying knowledge obtained in the human genome project.

Collins said last week that seven senior investigators — including two from his old base at the University of Michigan School of Medicine — have agreed to begin working at NIH in September and that he expects their number to increase to 20, directing an overall staff of 200, by 1995. NIH director Bernadine Healy says that such a wave of new talent would be "a critical example of the reversal of the brain drain from NIH". The new centre will broaden the existing \$100-million-a-year extramural gene-mapping effort at NIH (the Department of Energy will this year spend an additional \$62 million on the project) into DNA diagnostics, gene therapy and the search for disease genes.

President Bill Clinton has requested an increase of \$25 million in the fiscal year that begins on 1 October 1993 for the centre's intramural programme, an exceptional rate of growth when compared to the increase of 3.3 per cent requested for NIH as a whole. Even so, it is less than Collins had hoped for. He says that considerably more will be needed in fiscal year 1995 for "the critical mass we are hoping to achieve".

Achieving such growth will not be easy, however. Although Healy has been a strong proponent of setting up an intramural programme, she will be leaving on 30 June. If Congress does not grant NIH its wish, her successor will have to tax the other institutes to give Collins what he wants.

Collins, noted for his work in identifying genes for cystic fibrosis and, most recently, Huntington's disease, says he plans to continue spending at least half of his time at NIH doing science despite his administrative responsibilities as director of the genome centre.

C. M.

Clinton budget puts squeeze on accelerator labs

Washington. US high-energy physicists face another year of tight operating budgets, low capital spending and falling utilization of existing facilities, according to the budget for fiscal year 1994 (beginning on 1 October 1993) proposed last week by President Bill Clinton.

Although the budget for the Department of Energy would increase research spending in the field by 5.5 per cent, to \$627.8 million, the increase will be consumed by a doubling in major construction spending from \$43 million to \$86 million — meaning cuts in funding for operations and new equipment at five national laboratories.

Members of the department's High Energy Physics Advisory Panel (HEPAP) learned last week of reductions ranging from 11 per cent at the Stanford Linear Accelerator Center (SLAC) in California to 1 per cent in the \$102 million to be spent on university-based research. Four laboratories will each see reductions of around 2 per cent.

HEPAP will now ask the department to make revisions in health, safety and environmental regulations that cost the high-energy physics laboratories tens of millions of dollars a year to implement. It will also ask for a firm completion date for the Superconducting Super Collider (SSC) in Texas, which under the president's budget request would be deferred from 1999 to around 2003.

The physicists also want the Department of Energy and the National Science Foundation to complete quickly their review of whether to build the proposed 'B-meson factory' — an accelerator for which Clinton has requested \$36 million in 1994 in initial funding — at SLAC or at Cornell University in Ithaca, New York. The fear is that Congress will not appropriate the money this year unless a site has been chosen.

Restrictions on operating budgets have led to some expensive facilities being in experimental use for as little as eight weeks a year. Although several panel members complain that the discipline is now living off investments made in the 1980s, others question the wisdom of additional facilities such as the B-factory when the resources are lacking to operate existing ones.

The panel does not intend to review the work done last year by a panel chaired by Michael Witherell of the University of California at Santa Barbara, which set priorities for the field. But members are unhappy that nothing has happened so far on the Witherell recommendations and that political considerations have caused postponement of any decision on which facilities to curtail or close.

Colin Macilwain

US electronic highway needs consensus more than dollars

Washington. US vice president Al Gore has often likened his vision of a computer 'superhighway' to the interstate highway network built in the 1950s and 1960s for its potential to change the face of the country. But unlike that massive civil engineering project, Gore's plan to create a nationwide, high-capacity telecommunications network linking every home and office will require the federal government to first teach people how to drive and to set down the rules of the road.

The starting point for the administration's march towards an electronic super-highway is the high-speed networks now being tested at universities and government laboratories. One early is the transformation of NSFNET, supervised by the National Science Foundation, into a faster and more powerful research tool.

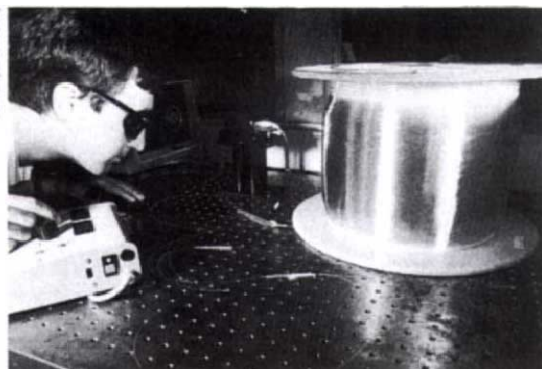
NSF plans to end its \$10-million support for the five-year-old, 45-megabit capacity NSFNET backbone and instead divert the money to its 20 regional networks to link up via the commercial alternatives. At the same time, it will soon ask for bids for a more powerful link, of at least 155-megabit capacity, between five large supercomputing centres that need the extra capacity and speed, in the process laying a foundation for the next-generation national network. But these plans depend upon a \$7-million addition this year to the programme's \$30-million budget that is part of President Bill Clinton's \$16-billion economic stimulus package, an initiative held up by Republicans in the US Senate.

Regardless of the fate of that proposal, the outline of the generation that will succeed the 155-megabit network is already visible on the horizon: under the auspices of the National Research and Education Network (NREN), networks with capacities from 600 megabits to 2.5 gigabits are to be tested at five sites across the country. The government-wide NREN programme, which was created by the High Performance Computing and Communications Act of 1991 championed by Gore, expects to spend \$400 million in the next five years.

In the meantime, Stephen Wolff of NSF is keen to see schools, libraries and other 'ordinary' users plug in to the existing 45-megabit network. "The network needs to have a robust presence in each community, insulated from policy and funding fluctuations", he says.

A bill (S.4) introduced by Senator Ernest Hollings (Democrat, South Carolina) and supported by the Clinton administration

would set up a national "information infrastructure programme" and would authorize several federal agencies to test high-speed network links to schools, libraries, hospitals and laboratories. An amendment would authorize extra money to pay schools and libraries to link up with NSFNET. Although some version of the Hollings bill — an overlapping and somewhat competing proposal (S.473) has been introduced by



How far ahead is an optical fibre network?

Senator Bennett Johnston (Democrat, Louisiana) — is likely to be passed, there are no plans to spend federal money on building any such network.

But the physical network is perhaps the easiest part of the jigsaw. "It is the software, applications, training, databases and getting people online that will be difficult", says Fiona Branton of the Computer Systems Policy Project, a computer industry lobby group.

The greatest impact of Clinton and Gore's strong political commitment to an electronic superhighway is likely to be on fixing the web of regulation that constrains private investment. The administration is in the process of appointing a task force on information infrastructure that would recommend a stable regulatory environment for high-speed telecommunications links. The computer industry strongly supports this proposal, believing that standard policies and protocols will give it a uniform home market for new products.

One important set of regulatory decisions will determine whether it makes commercial sense for US telecommunications companies to offer their customers services based on ISDN (Integrated Services Digital Network) in the near future. ISDN, which operates over copper wires, would serve as a stop-gap measure offering some digital communications. Its use may delay the immensely expensive construction of direct fibre-optic links to every home.

Colin Macilwain

NEWS IN BRIEF

New Delhi. Nine officials from the Indian subsidiary of the US Union Carbide Corporation will go on trial next month in connection with the release of methyl isocyanate gas from the company's Bhopal pesticides plant in December 1984 that killed more than 3,000 people and injured an estimated 200,000. Three others — including the former chairman of Union Carbide, Warren Anderson — face a separate criminal case.

The charges, brought against senior company officials as well as technicians on site during the disaster, include culpable homicide and voluntarily causing grievous hurt. The company, Union Carbide India Ltd, has pleaded not guilty and plans to appeal against a recent ruling by a Bhopal court that company officials were guilty of criminal negligence. In February 1989 the Indian government accepted compensation of \$470 million in return for dropping further prosecution, but last year the Indian supreme court revoked that stipulation.

K.S.J.

London. The UK Medical Research Council is seeking bids for a small number of centres of excellence for research into gene therapy, with a guarantee of five years funding for applicants who can demonstrate an identified clinical goal, a coordinated approach spanning gene manipulation, gene delivery and cell biology, and clinical protocols that include detailed evaluation of the psychological and social aspects of gene therapy.

The new centres are expected to be built around the work of existing research groups, with funding used to recruit sufficient extra staff to provide what the council describes as a "critical and balanced mass of research". The centres will be financed with money recently awarded to the MRC by the government for promoting a genetic approach to human health.

D.D.

Tokyo. Japanese and European X-ray satellites have observed X-rays from a supernova detected on 28 March, continuing a lucky streak by Japanese space scientists. The observations are only the second from a supernova and the earliest ever, and both have occurred shortly after the launch of a Japanese X-ray satellite (see page 585).

In March 1987, Ginga was launched a few weeks before the discovery of SN1987A in the Large Magellanic Cloud; a few months later it recorded a burst of X-rays from the supernova. In February, Asuka was launched and last week detected very strong X-ray radiation from the new supernova. The X-rays, in the range of 0.5–10 KeV, are believed to have appeared sooner this time because of the composition of the new supernova. It is thought to be a dense red super giant in contrast to the 1987 event, which resulted from the explosion of a less dense blue super giant.

D.S.

Religion in the genes

SIR — Notwithstanding the opinions expressed by Stuart Sutherland¹, I believe that religion can only gain from being investigated scientifically. However, any such investigations should be well-informed about religion as well as being scientifically valid. The discussion that follows attempts to treat the subject more meaningfully than do such analyses as those of Dawkins².

Any scientific study of religion should, in the first instance, take account of the fact that a central theme of religion (pathological variants excluded) is the attempt to maximize human 'goodness' (a quality I shall not attempt to define precisely). I speculate that religious practices have in part a genetic basis, involving genes linked to the potential for goodness. Societies in which this potential is actualized in a sizeable proportion of its members will tend to function more harmoniously and more efficiently, so that natural selection will tend to favour the presence in human societies of genes of this type.

Religious cultures revolve around the discovery that there are certain subtle kinds of experience that tend to foster the unfolding of the potential for goodness, and this constitutes the principal factor differentiating religion from humanism (the latter shares the general goals of religion, but adopts a primarily non-experimental approach towards their achievement).

Different religious cultures use different experiences, and different ideas, in pursuit of their common ends. Some, for example, take divinity as a central theme (experimentally as well as intellectually) while others (Buddhism for example) focus instead on qualities such as compassion and wisdom, and on seeing reality clearly. In time, religious cultures may lose sight of their original vision, and develop in malignant ways. The diversity of religion, and the possibility of its development in malignant forms, no more unconditionally discredits religion than do the diversity of languages, and the possible abuse of language, in the case of language.

Dawkins² criticizes religion on the grounds of apparent conflicts between religious beliefs and scientifically established facts. Convincing though such arguments may be at first sight appear to be, they are invalid. Consider the everyday situation of looking at a two-dimensional photograph, and perceiving it as a three-dimensional scene. Here, what I have an experience of, and what science says I am looking at, are two very different things. Thus the function of perception is not invariably that of objective representation; however, this

neither diminishes it nor deters scientific investigation. With religion, focusing on the factuality or otherwise of religious belief similarly misses the point: the significant questions in this context relate to the *functions* and *fruitfulness* of religious belief.

Emotive attitudes to religion have, it would seem, led to neglect of the important distinctions between science and expression of opinion in writings on the subject. I strongly urge the adoption of a more objective and insightful approach.

Brian D. Josephson

Cavendish Laboratory, Madingley Road, Cambridge CB3 0HED, UK

1. Sutherland, S. *Nature* **361**, 392 (1993).

2. Dawkins, R. *New Statesman and Society* **233**(5), 42–5 (18 December 1992).

Exchanging words

SIR — In his discussion of complexity theory (*Nature* **361**, 507; 1993), Ian Stewart states: "It is time science re-learned how to look outwards as well as inwards, to think about meaning as well as counting information, and to appreciate nature's semantics as well as its syntax."

I object to the use of syntax as an example of reductionistic thinking. If anything, syntax is the most complex property of language, even an emergent property, which might well be studied as a phenomenon "at the edge of chaos". Syntax is the very engine of language, the feature that allows virtually unlimited "cognitive compression," as Patricia Churchland has aptly described it. Syntax refers to word arrangement, made possible by grammar and in turn permitting nested structures and reflexivity.

Perhaps Stewart could change his sentence to read: we should appreciate nature's *syntax* as well as its grammar.

Gregory G. Dimijian

3277 Brincrest Drive, Dallas, Texas 75234–6568, USA

Interferon-tau

SIR — Ian Tizzard (*Nature* **361**, 676; 1993) is not alone in his puzzlement over interferon nomenclature. However, he is incorrect in his assumption that interferon-tau (INF- τ) had their origins as a typographical error, namely as the result of the inadvertent substitution of τ for γ on a word processor. Indeed, the IFN- τ are real entities whose existence was sanctioned by the International Committee on Interferon Nomenclature at its annual meeting in Toronto last

September. They are a group of type I IFN with clear sequence similarity to the related IFN- α , - β and - ω subtypes and were initially discovered as the result of the molecular cloning of their cDNA from trophoblast tissue of pre-implantation ovine embryos¹. At that time, they were regarded as curious variants of the IFN- ω (in 1987 themselves known as IFN- α_{II}), with which they share about 70 per cent sequence identity. However, it is now clear that the IFN- τ represent a separate multi-gene family and probably arose from the IFN- ω around 60 million years ago², at about the time the ruminants themselves began to diverge from other hoofed mammals.

Like other IFN types, the IFN- τ have potent antiviral and antiproliferative activities, but are unusual in that they do not themselves appear to be inducible by virus. As far as can be determined, their expression is limited to a few days of early embryo development when they play a role in the phenomenon of maternal recognition of pregnancy. Their discovery provided a new and completely unsuspected role for a type I IFN in reproductive physiology, and a problem in nomenclature that has taken several years to sort out.

R. Michael Roberts

Departments of Animal

Sciences & Biochemistry,

University of Missouri,

Columbia, Missouri 65211, USA

1. Imakawa, K. et al. *Nature* **330**, 377–379 (1987).

2. Roberts, R.M., Cross, J.C. & Leaman, D.W. *Endocr. Rev.* **13**, 432–452 (1992).

Geological Survey

SIR — I should like to add to what you say (*Nature* **361**, 99 & 101; 1992) about the British Geological Survey (BGS). First, the BGS hydrocarbons unit does in fact collect both offshore and onshore data submitted by companies under the terms of their licences.

Second, the BGS is responsible for making, and keeping up to date, geological maps and other interpretative publications on both onshore and offshore geology. For this purpose it needs access to all subsurface data as they become available for use. To put some of this information under someone else's control will make it more difficult for BGS to do its work efficiently. The strategic importance of this part of BGS activity has long been recognized, and was most recently confirmed by the report of Sir Clifford Butler's study group, published in 1987.

D. T. Donovan

Department of Geological Sciences,

University College London,

Gower Street, London WC1E 6BT, UK

Science and wealth creation

Mark Richmond

How can governments ensure that wealth is created from the scientific enterprise? In an abbreviated version of his talk at *Nature's* recent meeting¹, Sir Mark Richmond assesses the probable British strategy.

WITHOUT an excellent basic research base, I do not believe Britain can ever produce a highly developed workforce for the needs of its industry in the next century. Research provides the essential background against which the specific programmes of industry and commerce can be pursued. But even though a strong basic research base is essential, it is folly, in my view, to expect miraculous benefits to flow simply from having one, however distinguished. One must have some overall strategy and the necessary mechanisms to facilitate its exploitation.

The development of a 'mission' for the science base in general and for basic research in particular^{1,2} immediately raises issues of who is to shape and coordinate it. I would see the research councils (or analogous bodies), with their ability to fund research both proactively, in the context of a strategy, and responsively, responding to bids arising ultimately from the imagination of gifted scientists, as of crucial importance in this respect. Such a dual mode of proceeding allows a degree both of direction and of free thought to flow in science.

Of course, not all public money for research in Britain flows through the research councils. Departments of government have science budgets to pursue their objectives as set by parliament when they receive their parliamentary votes; any national strategy for wealth creation through research will need to integrate spending by these departments with that of the research councils.

Recent events have tended to work in the opposite direction. Up to the last general election, research in higher education institutions was supported very substantially by money flowing through a single department of government, the Department of Education and Science. Until then the two legs of the dual-support system for university research were the responsibility of one secretary of state. Now, with the setting up of the Office of Science and Technology (OST) and the demise of the University and Polytechnic funding councils to form regional funding councils, the dual-support system is in the hands of five government departments and five cabinet ministers. Some feel that this will give the consideration of science by the cabinet greater weight, others are not so sure. At all events, all the signs are that the integration of an overall wealth-

creation policy will not be helped by these changes. The birth of OST has had the effect of constructing additional interfaces without reducing the number elsewhere.

Then there is the matter of the money the research councils/OST spend compared with government spending departments. This is an issue of absolutely central importance to any strategy for the science base and for wealth creation. In fact so important is it that one can only imagine that the critical decisions in this area must have already been taken by the government before the announcement of the forthcoming White Paper (policy document) on scientific research. In one important regard matters in this area are obscure, at least to outsiders. It relates to how government departments bid for science money and how overall priorities are set. Each government department certainly bids for money under the annual public spending procedure, but it is quite unclear to those outside "the ring fence" whether these bids contain specific lines for research. The whole area of public expenditure bidding is covered by low cloud. Consequently it is unclear whether the government has anything approaching an integrated research budget — even in concept — or whether there are merely a series of individual departmental bids against headings of 'research' over which the chief scientist at OST casts an eye.

Perhaps one example will illustrate the importance of this issue. The pursuit of wealth creation will require an effective interface between the government's science spending and that of industry and commerce. As far as physics-based industries are concerned, the Department of Trade and Industry has a central role. Other government departments are in the lead for other areas: for example the Ministry of Agriculture, Fisheries and Food for the food industry and the Department of Health for pharmaceuticals. So who is to coordinate these activities, let alone coordinate them with the relevant research councils, who themselves are supporting (and in some cases doing) research in cognate areas?

The setting up of OST and the announcement of the White Paper with OST central to its production suggests that it is that department which is to play the coordinating role. But can one expect OST, itself a vote holder from

parliament for a part of the science base; to combine an executive role with respect to that money with an advisory/supervisory one for the spending of other government departments, particularly when they are likely to be bidding against one another? I do not believe OST, at least as operating at the moment, can be both executive and advisory/supervisory on the same topic.

So should all the spending departments' science money be transferred to OST? Probably. But unless things are already decided, there is not much time to arrange that, even in principle, before publication of the White Paper. Nor can one believe this is actually to happen — on a large scale at least. The alternative possibility — to transfer all OST's money to spending departments — would also have serious disadvantages. Apart from the politically difficult step of reversing the recent dual-support transfer, depriving OST of virtually all its funds would weaken it disastrously.

Because I believe the generation of advice should be clearly separated from executive control, OST may well have to be divided functionally to reflect two distinct roles. I don't feel scientists, or even those in industries concerned with wealth creation, should be too sanguine about this. The motivation for any decision in this area is likely to be almost exclusively one of raw politics, and the decision will have a profound effect on how science is managed in the future.

As well as influencing the structures and missions of the research councils, the White Paper may have a profound impact on the way they operate. Speaking personally, I am attracted to a 'mission-oriented' way of operating; but it would have important implications. In particular a research council would seek to fulfil its mission wherever the work could be done most effectively. For the Science and Engineering Research Council this would not necessarily be in higher education institutions, and so the White Paper could indirectly signal a sharp change in funding patterns. □

Sir Mark Richmond is chairman of the Science and Engineering Research Council, Swindon SN2 1ET, UK. In this article he is writing in a personal capacity.

1. Dickson, D. *Nature* **362**, 285–286 (1993).
2. Dickson, D. *Nature* **360**, 705–706 (1992).

Supernova brightens the horizon

Astronomy, as opportunistic a discipline as can be imagined, has been electrified by the lucky appearance of a supernova in a nearby galaxy. The spirit of 1987 is reborn.

HUNTING supernovae requires the patience of a bird watcher in the Sahara. Although one or two may be discovered in any week, most are in distant galaxies and so are hard to observe. The circulars ('telegrams') of the International Astronomical Union (now distributed electronically) are full of announcements of discoveries of supernovae in anonymous or obscure galaxies. Despite their enormous intrinsic luminosity, outshining the Sun a billionfold, they barely register in astronomical images because they are so far away.

Another difficulty is that supernovae, despite their immense energy — or perhaps because of it — are over almost before they start. After countless millions of years of unrelieved obscurity, a star will exhaust its nuclear fuel, collapse and then explode in a matter of seconds. What we see is merely the aftermath as the excess heat from the remnant is radiated away over the subsequent years. Andy Warhol's 15 minutes of fame is more than most stars can aspire to.

But patience pays off. The appearance of supernova 1987A (SN1987A) a little over six years ago, although it was outside the Galaxy, showed how. That was the first nearby supernova that could be seen well since 1604; it was the first opportunity to confirm by observation theories of how stars explode. The detection of neutrinos from the supernova, part of that confirmation, inaugurated the field of neutrino astronomy (the solar version excepted) and also served as an unprecedented test of ideas about these ghostly particles.

Supernova 1987A also proved the value of the IAU circulars. Within a day, telescopes all over the Southern Hemisphere and also aboard astronomical satellites were trained on the Large Magellanic Cloud, with the data broadcast daily to astronomers. Something of that heady atmosphere has been recaptured in the past fortnight, with the discovery by Francisco Garcia, a Spanish amateur astronomer, of a supernova, SN1993J (the tenth recorded this year), in the spiral galaxy M81, in the constellation of the Great Bear (*Ursa major*).

Except for SN1987A, this is the brightest supernova at the Earth in more than 20, perhaps 50 years. SN1993J occurred in a galaxy that, like the Large Magellanic Cloud (host to SN1987A), is among the most closely watched in the sky: the 'M' in its name shows that it is conspicuous enough to have been included in the catalogue compiled by Charles Messier more than 200 years ago.

One of the immediate benefits was that old astronomical plates could be searched to

find out which star had exploded. The supernova's field is crowded, but researchers are pretty satisfied that they have found the progenitor; ultraviolet observations in a few weeks should confirm that (or otherwise). Arguments about the progenitor of SN1987A rumbled on for much longer.

As most theories require, the progenitor was a red supergiant star, the typical inflated state of stars that have exhausted the hydrogen in their cores and are burning heavier elements. The progenitor was not as red as astrophysicists would really like, but neither was it as abnormal as the progenitor of SN1987A, which was a blue supergiant. Nor was it as difficult to identify, even though M81 is 70 times further away than the Large Magellanic Cloud. The condition of a star on the point of exploding is evidently not strictly constrained.

The popularity of M81 with astronomers also meant that photographs of this region of sky were available within hours before and after the supernova exploded. At about midnight on 27 March, there was no sign of a supernova on an image obtained by Jean-Claude Merlin, an amateur French astronomer; 9 hours later, it was there on an electronic image obtained by Bill Neely, another amateur, in New Mexico. Only rarely can astronomers pin down the moment of eruption so accurately. SN1987A was unique, of course, in that the neutrino pulse, produced at the precise moment of explosion, gave a precise time zero.

SN1993J has nicely conformed to the prescribed pattern of so-called type II supernovae. After the exhausted stellar core, no longer sustained by nuclear reactions, collapsed and then rebounded, a shock wave was driven out through the stellar envelope, raising the stellar surface to white heat (10^5 – 10^6 K) within a few hours. This is what the astronomers first saw. Over the next day or so, the star's brilliance increased to the point that it could be seen with a small telescope or even binoculars (10th magnitude).

The star soon started to fade as the old stellar surface started to cool. But as the shock wave ran into the tenuous wind that had been radiated by the progenitor star in its declining years, X-rays were detected on 3 April by Rosat, the astronomical satellite launched in 1990, and subsequently by the new Japanese satellite, ASCA. These betoken gas heated to 10^8 K or more. As material in front of the shock wave thinned out, radio waves from the shock could escape, to be detected on 5 April at the Mullard Radio Astronomy Observatory in Cambridge.

So SN1993J has now appeared over almost the full extent of the electromagnetic spectrum, although neutrinos have not been seen (unsurprisingly given the 3.2 megaparsec distance to M81). One day, gravitational waves from supernovae may be detected, but the appropriate antennae have yet to be built — the last frontier in observational astronomy has yet to be crossed. Curiously, the one player missing from the field is the Hubble Space Telescope: while amateurs and large projects alike have turned towards M81, Hubble, beset by technical problems (a computer glitch is hindering operations) and encumbered by a ponderous bureaucracy, has failed so far to take advantage of this 'target of opportunity'.

Even so, the interest in SN1993J involves a certain clutching at straws. What astronomers really need is to see a supernova in the Galaxy. Current estimates are that one appears every 40–50 years. But the last two to be seen appeared in 1604 and 1572. Before that, there were visible supernovae in 1006 and 1054. Unless the estimates are wrong, the others must have been hidden by the lanes of dust and gas that criss-cross the galactic disk.

What accurate observations beyond the Galaxy are showing is that the characteristics of type II supernovae are remarkably variable. All seem to be 'atypical', which begs the question: what is typical? Zwicky's inspired guess 60 years ago of the basic mechanism has proved to be remarkably accurate — right down to the prediction of neutron stars — but the details are turning out to be far more exciting. SN1987A has yet to reveal its neutron star, for example.

The other regret has been that nobody has yet seen a nearby type I supernova. Far brighter than type IIs, these explosions arise in white dwarf stars, possibly triggered by the accretion of a critical amount of mass. The preconditions seem so precisely tuned that many think that all type Is (or more specifically type Ias) have to be alike. For cosmologists, this offers the prospect of reliable beacons ('standard candles') to be seen far across the Universe, and a means to measure the cosmic distance scale and Hubble's constant, which gives the rate of cosmic expansion. Indeed, the catalogues are stuffed with the parameters of distant type I supernovae. The trouble is that none has been near enough to give an unequivocal calibration of the scale. As one astronomer remarked ruefully, had SN1993J been a type I, we would have known within days the answer to the most burning astronomical question.

Roland Pease

Instant teleportation

Tony Sudbery

THE latest idea from science fiction to be adopted by quantum physicists (following parallel universes¹ and time machines²) is teleportation, discussed in a paper which has just appeared in *Physical Review Letters*, by Charles Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres and William Wootters³. The idea behind teleportation is that a physical object is equivalent to the information needed to construct it; the object can therefore be transported by transmitting the information along any conventional channel of telecommunication, the receiver using the information to reconstruct the object.

In classical physics there is nothing very implausible about this; indeed, such devices already exist (for the surfaces of objects, and for detail on the scale of millimetres) in the form of fax machines. These even embody a more extreme form of the science-fictional idea by producing a *copy* at the point where the message is received and leaving the original at the point of transmission. However, the principle of the fax machine can operate only in the realm of classical physics; if one steadily reduces the scale at which it faithfully reproduces detail, one will eventually run foul of the uncertainty principle of quantum mechanics in one form or another, with the consequence that the object in its original state is bound to be destroyed by the process of scanning for transmission. This was proved by Wootters and Zurek some ten years ago⁴ under the slogan "A single quantum cannot be cloned".

Those familiar with the peculiarities of quantum mechanics might suspect that there are further obstacles which will make it impossible to transmit a microscopically complete description of an object. According to conventional quantum theory, any attempt to gain information about an object nearly always changes its state, so that the information gained is true only after the object has been examined, not before. There are, it is true, special circumstances in which examining the object does not change it, but one can never tell if this was so purely on the basis of the result of the examination. Empirical information about a quantum object is always ambiguous, and therefore insufficient to reliably reconstruct a previously unknown object.

These general principles remain unchallenged, but they are put in an interesting new light by the idea of quantum teleportation proposed by Bennett and colleagues. The authors divide the complete description of the quantum

object into a classical part, consisting of information obtained by an experiment which can be transmitted or broadcast on any conventional channel, and the quantum residue, undetectable and, one would have thought, untransmittable except by sending the object itself. But teleportation provides another way of



transmitting quantum information, using the concept of 'entanglement' of two objects. This is the situation whose peculiarities were first pointed out by Einstein, Podolsky and Rosen (though the word is due to Schrödinger) in which the properties of the two objects are correlated in such a way that an observation of one of them appears to affect the other — instantaneously.

In the device of Bennett *et al.* the object to be described (call it X) is entangled with another object A by a quantum measurement of a joint property of the two particles. Normally this measurement would destroy the full quantum information about X . However, if A is already entangled with a third object B , the effect of the measurement on X and A is to transfer the state of X to B — not faithfully, but after a transformation which depends on the result of the measurement. The transformation can then be undone by someone who knows the result of the measurement, leaving B in the state in which X started. It has become the teleported version of X .

Thus when the device is used by a sender Alice and a receiver Bob, the sequence of events is as follows. When it is set up, Alice and Bob are provided with an entangled pair of objects, A and B . When Alice wants to send an object X to Bob, she performs a joint measure-

ment on X and A . Bob's object B immediately becomes one of the transformed versions of X . Alice then informs Bob of the result of her measurement, using a classical channel of communication; this tells Bob which of the transformed versions he has, so that he knows which method to use to undo the transformation and obtain his copy of X .

This method of sending the information about X has the advantage over simply sending X itself that Alice does not need to know where Bob is. Once the device has been set up by entangling A and B , Bob can wander around with B without breaking the channel for the lone copy of the quantum information. The classical information, which is reproducible, can be broadcast so that Bob can receive it wherever he is.

Whether or not this device can be realized in practice, it may, as a thought experiment, be relevant to some unresolved issues in the interpretation of quantum mechanics, though serving more to illustrate the issues than to resolve them. First, is there any genuine instantaneous influence or transmission in the Einstein-Podolsky-Rosen (EPR) experiment? Second, does the quantum state vector (or wavefunction) describe objective physical properties of an object? The two questions are closely related. The EPR experiment is incorporated in the teleportation device as the first part of the operation, which I have described as an instantaneous effect of Alice's measurement on Bob's object. However, it is only the abstract quantum description of Bob's object, its state vector, that is instantaneously affected; this can have no consequences observable by Bob until he receives the results of Alice's measurement.

Thus the transmission of observable effects is subject to the same restrictions as classical communication, and there is no action at a distance — Einstein, Podolsky and Rosen's bugbear — unless one believes that the state vector has real physical existence. It is not universally accepted that it does. Recently, Yakir Aharonov, together with Jeeva Anandan⁵ and Lev Vaidman⁶, has proposed methods to observe the full details of the state vector, converting it all to classical information and leaving no quantum residue; but these observations take a long time and work only when one knows in advance what one

1. DeWitt, B. S. & Graham, N. (eds) *The Many-Worlds Interpretation of Quantum Mechanics* (Princeton University Press, 1973).

2. Aharonov, Y., Anandan, J., Popescu, S. & Vaidman, L. *Phys. Rev. Lett.* **64**, 2965–2968 (1990).

3. Bennett, C. H. *et al.* *Phys. Rev. Lett.* **70**, 1895–1899 (1993).

4. Wootters, W. K. & Zurek, W. H. *Nature* **299**, 802–803 (1982).

5. Aharonov, Y. & Anandan, J. *Phys. Rev. A* (in the press).

6. Aharonov, Y. & Vaidman, L. *Measurement of the Schrödinger Wave of a Single Particle* (Tel Aviv preprint, 1993).

is going to observe.

Nevertheless, the teleportation device might be taken as a reason to believe in the reality of the quantum state vector; if something can be transmitted from one object to another, it might be argued, it must be real. The debate will continue. Meanwhile, the idea of quantum tele-

portation possibly represents a further step for Einstein, Podolsky and Rosen on the road from pure philosophy to practical application. □

Tony Sudbery is in the Department of Mathematics, University of York, Heslington, York YO1 5DD, UK.

MOLECULAR GENETICS

Colour-coded switches

Ian J. Jackson

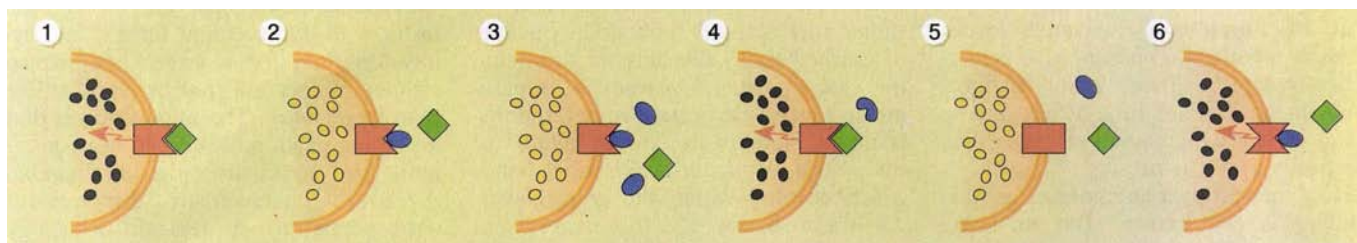
MOLECULAR analysis of mouse coat-colour genetics has revealed a great deal about all kinds of biological phenomena. The most striking recent examples are the classical *W* mutations, which turn out to be in the receptor tyrosine-kinase gene *Kit*, and the *Steel* mutations in its ligand. Work on two more classical coat-

product results in black pigment only, overproduction in yellow only, and the wild-type pattern must be due to a pulse of product midway through hair growth.

The complication comes with the *e* locus. At this locus, which acts autonomously within the melanocyte, the recessive mutations produce yellow pigment,

taneously produced new alleles of *a* and, at the linked locus, *limb deformity* (*ld*; ref. 5). The inversion, it seemed, interrupted both genes. As *ld* had previously been cloned, it was a relatively simple matter for Rick Woychik's group (who had previously participated in cloning *ld*) to step across the inversion into *a*. The inversion breakpoint lies within the third intron of an 18-kilobase gene, encoding a small 131-amino-acid protein, which appears to be a secreted protein having a signal sequence, but no transmembrane domain^{1,2}. The gene has the characteristics one might expect. It is expressed normally only in skin, but not in melanocytes, in a pulse between days 2 and 7 after birth — just at the time when the melanocytes switch to making the yellow band on the hair.

The *a* mutations are also as one might predict. An extreme recessive allele contains an intragenic deletion. The classical *non-agouti* mutation, which results in



The molecular biology of black or yellow coat colour. 1, In wild-type mice, binding of α -MSH (green) to the product of *extension* (E, red), is required for synthesis of black pigment; 2, α -MSH binding is antagonized by the product of *agouti* (A, blue) and results in yellow pigment. 3, The dominant mutation in *agouti* means constant expression of A and hence constant yellow pigment; 4, recessive mutations resulting in absent or non-functional A mean that α -MSH binds continuously and constant black pigment is expressed. 5, With a recessive mutation giving non-functional E, α -MSH does not bind and the consequence is constant yellow pigment; 6, dominant mutations of *extension* result in constitutive signalling, even in the presence of A, and hence constant black pigment.

colour genes, *agouti*^{1,2} and *extension*³, now shows that these are also part of a receptor/ligand complex, but with a twist. It looks as if mutations at the *agouti* locus affect not a ligand but an antagonist which prevents ligand (α -melanocyte-stimulating hormone) binding to its melanocyte-specific receptor, the product of the *extension* locus.

The underlying genetics are quite straightforward. Wild-type mouse hair is black, with a yellow stripe midway along the hair's length; this is the *agouti* pattern. Mutations at both *agouti* (*a*) and *extension* (*e*) affect this pattern. Recessive (presumably loss-of-function) *a* mutations result in loss of the yellow band and in more-or-less completely black animals. Dominant (gain-of-function) mutations at the same locus produce wholly yellow animals. Chimaera and grafting experiments show that the *a*-gene product does not act within the pigment-producing cell, the melanocyte, but exerts its influence from the surrounding microenvironment. So far so good. The *a* locus makes a protein that switches the melanocytes from black melanin synthesis to yellow. Loss of the

whereas the dominant alleles result in black. So the default pathway seems to be that of yellow pigment synthesis, and the simplest explanation of *e* encoding the *a* receptor cannot be correct (as in this case the loss of function of the receptor should give black pigment synthesis). The first clue to what is actually going on came many years ago when it was shown that injection of the melanocyte-stimulating hormone, α -MSH, promoted black pigment synthesis after injection into dominant *a* mutant yellow mice, but not in yellow recessive *e* mutant animals. Thus a model arose (see figure, and ref. 4 for review) in which *e* encodes the α -MSH receptor (MSH-R); α -MSH is required for black pigment synthesis; and *a* is an antagonist of α -MSH which prevents binding or signalling and results in yellow pigment.

The two genes have now been cloned, one by positional cloning^{1,2}, the other by examination of a candidate gene³. Although a number of laboratories had made long-range physical and genetic maps around the *a* gene, the key to its cloning was finding a radiation-induced chromosomal inversion, which simul-

some yellow pigment synthesis in a few parts of the body, has a large (more than 11 kilobase) insertion into the first intron. No transcript is detectable from the locus; presumably the insertion eliminates transcription in all but a few locations. Two dominant alleles have deregulated overexpression of *a*, which leads to completely yellow animals. One, *Viable yellow*, has overexpression of a normal-sized messenger RNA, whereas *Lethal yellow*, which is associated with a chromosomal rearrangement, overexpresses a different-sized mRNA. Both alleles have expression of the *a* product in all tissues, which gives rise to various effects, of which more below. Finally, the *black and tan* mutation results in mice that are black on their backs (where no *a* transcript is detectable) but yellow on their bellies (where constitutive overexpression is seen).

The melanocyte α -MSH receptor was cloned last year⁶. There is a family of distinct melanocortin receptors expressed differentially, but they were known to be G-protein-coupled and likely members of a much larger superfamily of seven-membrane-spanning receptor

molecules. Mountjoy *et al.* of Roger Cone's laboratory used the polymerase chain reaction with degenerate primers to amplify a sequence from a melanoma line expressing high levels of MSH-R. With this they obtained the full-length sequence, which they found encodes the functional receptor.

The cloned receptor made it possible to see if this receptor is the product of the *e* locus. Mapping the murine MSH-R gene showed that it was very close to, and therefore a candidate for *e*. The final piece of the story fell into place last month, enabling Cone and colleagues³ to determine the sequence and activities of the mutant alleles. The recessive allele has a frameshift mutation between transmembrane domains 4 and 5, which results in a non-functional receptor. So, no α -MSH signal is received and the melanocyte defaults to produce yellow pigment. Three dominant mutations have been characterized, and all are point mutations. One, *Sombre-3J* is constitutively activated to around 60 per cent of hormonally activated levels, whereas another, *Tobacco*, is a hyperactive receptor. These mutants signal even in the absence of α -MSH, and so result in black pigmentation even when the product of *a* is present.

How that product antagonizes α -MSH binding is not known. But an extra dimension to the story ensures that there will be keen interest in finding out. The dominant (yellow) *a* alleles have several effects, which include obesity, resulting in diabetes, and a propensity to develop tumours. So, ectopic expression of *a* leads to these phenomena. Is the MSH-R normally expressed elsewhere, and being abnormally blocked there by the *a* product in the mutants? Probably not, as the recessive *e* mice are not obese nor especially susceptible to tumours. More likely, ectopic *a* expression is interfering with other receptor/ligand complexes elsewhere. If the mutant effect on these complexes results in obesity, their identification may provide clues to the basis of some forms of obesity, and perhaps pointers to pharmacological treatments. Finding where these complexes are is a first step, and that is likely to be achieved using transgenic mice in which the *a* gene is driven with different tissue-specific promoters. □

Ian J. Jackson is in the MRC Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, UK.

No need to stick to the point

Charles G. Sammis

ONE of the primary goals of earthquake mechanics has been to understand the physical instability that nucleates an earthquake. The hope is that such an understanding will, one, guide the search for physical precursors that might have a predictive value; and, two, relate nucleation to fault structure, helping to identify likely nucleation sites of impending earthquakes. Mechanical models of fault-zone stability are complicated by the observation that natural faults can be geometrically complex and commonly contain a layer of crushed rock or 'fault gouge'. Elsewhere in this issue (*Nature* 362, 618–621; 1993), C. Marone and B. Kilgore present the first convincing link between the structure of such a fault zone and its frictional stability.

In a series of frictional sliding experiments on simulated fault zones (two sliding rock surfaces separated by a layer of crushed rock) the authors show that the characteristic displacement, a parameter which is important in determining frictional stability, is directly related to the width of shear-localization bands which develop within the gouge layer. As these bands are observable not only in laboratory experiments, but also in natural fault zones that have been exhumed by a combination of tectonic uplift and erosion, Marone and Kilgore's observations represent an important breakthrough in the long-standing problem of how frictional-stability studies in the laboratory should be scaled to model earthquake nucleation on natural faults.

The problem with scaling from the laboratory to the field is that it is not clear how parameters measured for rock-on-rock friction apply to zones of crushed rock. In the laboratory, studies of frictional sliding between rock surfaces have found a stick-slip instability from a velocity-weakening phenomenon, in which the coefficient of friction decreases with increased sliding velocity. Velocity weakening in rock can be understood in terms of the classical asperity model of Bowden and Tabor, in which surfaces are in contact only at microscopic bumps, the asperities.

In the asperity model, velocity weakening comes about because the asperity strength is time dependent. The longer an asperity contact is maintained, the stronger it becomes. At higher sliding velocities, the average lifetime of an asperity is shorter, it is weaker when it breaks and, therefore, the coefficient of friction is slightly lower. In this model, the characteristic displacement is related to the average distance between asperities, and can be thought of as the

displacement required to change a population of contacting asperities to a completely new population having strengths appropriate to the new sliding velocity. The characteristic displacement is therefore the sliding distance required to weaken the coefficient of friction following an increase in sliding velocity.

What is the significance of all this to fault stability? It turns out that the reduction of friction associated with an increase in slip speed and the critical distance over which it occurs define a rheological stiffness. The stick-slip instability requires that the stiffness of the driving system (the laboratory apparatus or the elastic continuum surrounding the fault) be less than this rheological stiffness. In the case of faulting, the stiffness is related to the size of the slipping dislocation patch. Driving stress falls faster with displacement for a small dislocation than for a larger one, hence smaller dislocation patches are stiffer than larger ones. The implication is that a slipping patch must exceed some minimum size to nucleate an earthquake. A larger characteristic displacement corresponds to a less-stiff rheology which requires a correspondingly larger patch for nucleation. If the characteristic displacement is large enough, it is conceivable that the motion on a fault could be accommodated by a series of creep events without the nucleation of earthquakes.

The characteristic displacement measured in the laboratory is of the order of a few micrometres for rock-on-rock friction. If this same value is relevant for faults, earthquakes will be nucleated from very small patches and the chance of observable precursory creep or other pre-failure phenomena is nil. However, as mentioned above, real faults in nature are not well modelled as two rock surfaces in contact. The observation of a layer of crushed rock separating the fault surfaces raises the question of what constitutes an asperity, and what physically determines the characteristic displacement in such systems. Previous laboratory studies of frictional stability in artificial gouge layers found that the characteristic displacement is, in fact, larger in such systems (10–100 micrometres), but no clear dependence on either layer thickness or particle size had been observed which would allow direct scaling from the laboratory to the field.

Marone and Kilgore's principal contribution is the design of a set of experiments that clearly separates the effects of layer thickness, particle size and shear strain on the characteristic displacement.

1. Bultman, S. J., Michaud, E. J. & Woychik, R. P. *Cell* **71**, 1195–1204 (1992).
2. Miller, M. W. *et al.* *Genes Dev.* (in the press).
3. Robbins, L. *et al.* *Cell* **72**, 827–834 (1993).
4. Jackson, I. J. *Bioessays* **13**, 439–446 (1991).
5. Bultman, S. J., Russell, L. B., Gutierrez-Espeleta, G. & Woychik, R. P. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8062–8066 (1991).
6. Mountjoy, K. G., Robbins, L. S., Moritud, M. T. & Cone, R. D. *Science* **257**, 1248–1251 (1992).

By also measuring subtle volume changes associated with changes in sliding velocity they argue, convincingly, that the characteristic displacement scales as the thickness of the gouge that actually participates in the deformation — that is, the width of the shear localization bands. Thus armed with a physical interpretation of the characteristic displacement in terms of observable fault-zone structures, the problem of scaling frictional-stability parameters measured

in the laboratory to the field can now be sensibly addressed. How such structures are related to more readily observable fault-zone parameters such as depth, width and total displacement has yet to be determined, but at least the question has been framed. □

Charles G. Sammis is in the Department of Geological Sciences, University of Southern California, Los Angeles, California 90089, USA.

PALAEONTOLOGY

Ground rules for early birds

Angela C. Milner

UNTIL very recently, our knowledge of the early evolution of birds was handicapped by the paucity of the Mesozoic fossil record and was restricted largely to the 147-million-year-old *Archaeopteryx* from the Upper Jurassic of Bavaria, Germany. But a trickle of discoveries, the most recent of which is reported by Perle *et al.* on page 623 of this issue¹, is now providing tantalizing clues to the complexities of the morphological transformation series of characters leading to the development of powered flight apparatus.

The new specimens, designated *Mononychus*, come from the late Upper Cretaceous of Mongolia (that is, about 75 million years ago). The creature is represented by three-dimensional remains, whereas most other key examples are essentially two-dimensional, crushed and often fragmentary, which makes anatomical interpretation difficult. So *Mononychus* provides a character set of great importance to understanding primitive birds; this set can be interpreted unambiguously and it indicates a transitional position between *Archaeopteryx* and all other birds. However, alongside its primitive bird characters, *Mononychus* possesses a uniquely derived forelimb: it is short and powerfully built, and the hand is reduced to a single digit bearing a robust claw. Clearly, this animal did not fly.

This unexpected find demonstrates how little is known about the radiation of birds in the Mesozoic. It is widely, but not universally, accepted that birds are descended from advanced theropod dinosaurs with a specialized wrist structure (maniraptorans²). *Archaeopteryx*, the earliest avialian (the taxon that includes it and all other birds³) possessed a wing structure identical in many respects to that of modern birds. Impressions of the wings and tail feathers are uniquely well-preserved in the fine-grained lithographic limestone matrix of the famous Solnhofen Limestones of Bavaria. The

primary and secondary flight-feather count in the wing is identical⁴ to that of modern birds, as is the structure of the flight feathers in which asymmetric vanes confer aerofoil characteristics⁵. As far as its skeletal anatomy is concerned, there are few characters that distinguish *Archaeopteryx* from maniraptoran theropod dinosaurs.

What of other recent finds? Lower Cretaceous birds from 135-million-year-old rocks in China (*Sinornis*⁶), and from very slightly younger deposits at Las Hoyas in Spain (*Iberomesornis*⁷, *Concornis*⁸), demonstrate a more derived pectoral condition than *Archaeopteryx* in that they have a strut-like coracoid, furcular process and ulna longer than humerus. All three also share a pygostyle (reduced tail formed from fused caudal vertebrae) for support of a tail feather fan, which is important in flight manoeuvring at slow speeds. *Sinornis* and *Concornis* are more derived than *Iberomesornis* in their possession of a keeled sternum; *Concornis* shares derived characters with modern birds, including fusion of elements in the manus and tibiotarsus. These Lower Cretaceous birds are small — *Iberomesornis* and *Sinornis* are the size of a house sparrow, *Concornis* about twice that — which suggests that a size-filter may have operated as a selection pressure in the development of powered flight. The acquisition of a suite of evolutionary novelties relating to flight and perching apparently preceded 'refinements' of other parts of the skeleton which shifted the centre of gravity and balance forwards towards the forelimbs.

The bird record from the Upper Cretaceous is sparse, but demonstrates the establishment of a range of advanced fliers, secondarily flightless seabirds (hesperornithiforms), a flightless cursorial (running) land bird *Patagopteryx*⁹, and representatives of two modern orders. *Mononychus* postdates most of those occurrences and is evidence of a

hitherto unknown radiation of primitive, thrush-sized cursorial birds. Perle *et al.* offer two equally parsimonious alternatives concerning the acquisition of flight. Either it was primitive for all avialians and lost in the lineage leading to *Mononychus*, or flight arose independently in *Archaeopteryx* and the derived sister group (Ornithothoraces¹) to *Mononychus*. The latter implies that an identical flight-feather structure and complement arose convergently. Because *Mononychus* had a keeled sternum¹, a derived character associated with flight which it shares with all other post-*Archaeopteryx* birds, it is perhaps more likely that it represents a lineage that diverged in late Jurassic or early Cretaceous times and became secondarily flightless. A test of this hypothesis awaits the discovery of further metornithine¹ birds.

Feathers are rarely preserved in the fossil record. In the Lower Cretaceous their only associations with skeletal material are poorly preserved traces in *Concornis*, and the record is meagre in the Upper Cretaceous. However, it is a logical assumption that feathers were present in all birds derived with respect to *Archaeopteryx*, including *Mononychus*, for insulation, if not for flight. Developmentally, feathers and reptilian scales are homologous structures; down feathers must have arisen, initially as means of insulation, in the lineage of maniraptoran theropod dinosaurs from which birds arose. Another find from the Upper Cretaceous of Mongolia, a small theropod dinosaur, *Archaeornithoides*¹⁰, raises the possibility, as has been suggested in some cladistic schemes, that the bird lineage branched off the theropod line before the advent of a range of advanced dinosaurs. Feathers may, therefore, have been widespread among bipedal carnivorous dinosaurs as an insulating outer layer, as well as being universal in birds. □

Angela Milner is in the Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.

1. Perle, A., Norell, M. A., Chiappe, L. M. & Clark, J. M. *Nature* **362**, 623–626 (1993).
2. Schultze, H.-P. & Trueb, L. (eds) *Origins of the Higher Groups of Tetrapods* (Comstock, Ithaca, 1991).
3. Gauthier, J. *Mem. Calif. Acad. Sci.* **8**, 1–56 (1986).
4. Heimann, G. *The Origin of Birds* (Witherby, London, 1926).
5. Feduccia, A. & Tordoff, H. B. *Science* **203**, 1021 (1979).
6. Sereno, P. C. & Rao, C. *Science* **255**, 845 (1992).
7. Sanz, J. L. & Bonaparte, J. F. in *Papers in Avian Paleontology Honoring Pierce Brodkorb* (ed. Campbell, K. E. Jr) 39–49 (Nat. Hist. Mus. Los Angeles, Sci. Ser. 36, Los Angeles, 1992).
8. Sanz, J. L. & Buscalioni, A. D. *Palaeontology* **35**, 829–845 (1992).
9. Alvarenga, H. M. F. & Bonaparte, J. F. in *Papers in Avian Paleontology Honoring Pierce Brodkorb* (ed. Campbell, K. E. Jr) 51–64 (Nat. Hist. Mus. Los Angeles, Sci. Ser. 36, Los Angeles, 1992).
10. Elzanowski, A. & Wellnhofer, P. *Nature* **359**, 821 (1992).

Organizing the cerebrum

Jack Price

THREE papers in this issue¹⁻³ address the subject of whether cellular migration in the embryonic forebrain is strictly radial or whether it also has a tangential component. Why should such an esoteric topic be of so much interest? The answer lies in the various theories developmental neurobiologists have constructed to explain how the cerebral cortex becomes patterned.

As primates, we tend to be particularly proud of the cerebral cortex because it is the brain region that distinguishes us from what we call lower vertebrates. The cortex is essentially a continuous sheet of tissue sitting like a cap over the rest of the forebrain, but its most remarkable feature is its organization into functional areas. This makes it less like a sheet and more like a patchwork quilt, each patch being dedicated to a particular function such as vision or touch.

Two broad theories have arisen to explain the development of this form of organization. One idea, the radial unit hypothesis, is that the patchiness arises from a patchiness already present in the ventricular zone (VZ), the layer of precursor cells that generates the cortical neurons and glia⁴. The alternative theory is that the organization is imposed on the cortex by later influences such as the ingrowth of fibres from other brain regions⁵. Recent results have shifted the majority opinion from the first of these theories towards the second (see ref. 6 for review), but there is evidence both ways. The way cells migrate is a crucial element in this discussion.

Guidance

Primarily, cell migration is important in neural development because cells are generated far from their adult position, so they need precise guidance from one site to the other. For cortical neurons, the path is primarily radial from the VZ to the cortical grey matter, which develops outside the VZ in concentric layers. This radial migration is undisputed. The issue is whether there is also appreciable tangential dispersion, which constitutes a test of the radial unit hypothesis because pattern in the VZ would tend to be disrupted by tangential dispersion.

The first real evidence for tangential dispersion came from experiments using retroviral lineage labelling (see ref. 1 for references). In this technique individual VZ cells are directly labelled, allowing the behaviour of their progeny to be observed. The important question, though, was not just whether cells dis-

perse but whether they disperse freely across cortical areas.

Subsequent experiments have left little doubt that some boundaries can be crossed, notably in the formation of the hippocampus⁷. Evidence relating to the neocortex now comes from the paper on page 632 of this issue¹, in which Walsh and Cepko apply their innovation⁸, a library of 100 viruses detected by histochemistry and the polymerase chain reaction, to this part of the rat brain. With this method, they can distinguish dispersed progeny of single, labelled cells however far they spread. As in their previous study⁸, they label VZ cells with virus early in corticogenesis (but at a lower infection density than previously), and allow development to proceed for 3, 6 or 10 days. They discover that tangential dispersion is relatively modest over the first 3 days, but much more marked for about half of the clones over the next 3, so much so that some clones apparently spread out of the cortex into neighbouring regions. After 6 days, the picture remains relatively stable, although one clone showed particularly dramatic dispersion after this point. Generally, there is less dispersion than in the previous study, but this could reflect the slightly younger animals used here.

This biphasic dispersion is not discussed in detail by the authors, but it suggests that tangential dispersion is confined to the period of gliogenesis and late neurogenesis. This result also provides an important control for the authors' earlier study. Their technique was criticized for underestimating the likelihood that widely dispersed clones could actually be multiple infections by the same virus type⁹. This can be discounted if one set of viral infections analysed at two time points gives dramatically different degrees of dispersion. That is what is reported here, although the conclusion that dispersion is limited after 3 days rests on the evidence of very few clones. Nonetheless, it seems that some clones do disperse considerably.

The unanswered question is which cell types are dispersing, especially because some clones do not disperse widely. Some cells, such as oligodendrocyte precursors, are known to spread widely, but are not likely to be important in cortical patterning. The neurons are what are important, as well as any cells, such as radial glial cells, that could conceivably play a role in defining cortical areas.

Does tangential dispersion mean that the radial unit hypothesis is dead? In its strongest form, probably yes, but there are several ways in which pattern could

be transmitted from the VZ to the cortex without a point-to-point specification obeyed by every cell. The possibility that only a subset of the cells is important in specification is suggested by the paper by Tan and Breen on page 638².

Mosaicism

These authors generated a mouse line carrying a *lacZ* transgene on an X chromosome. Because of X inactivation, female mice will express the gene in only half of their cells. Such mice will be mosaic, half of their cells being blue and half white. This mosaicism is established before corticogenesis, so the VZ is a mix of blue and white cells, with no apparent pattern. Tan and Breen show, however, that the cortices of adult animals have blue and white radial stripes. They are variable in position and width, suggesting that they do not represent functional boundaries within the cortex, but instead reflect the random mosaicism of the VZ. Nonetheless, the stripes themselves point to strict radial migration having taken place, for tangential migration would disrupt the striped pattern.

How can this be, given the retroviral results? The answer is that the stripes are not pure; there are blue cells in the white stripes and vice versa. In fact, a third of the cells throughout the width of each stripe are the wrong colour whether the stripes are thick or thin. This is good evidence for two populations of cells, one migrating strictly radially and preserving the stripes, the other dispersing tangentially. Walsh and Cepko's data suggest something similar, although the authors do not discuss this point.

If some VZ cells move strictly radially, then we potentially have a means for conserving positional information during migration. This is comforting because, as I have mentioned, there is evidence for some early specification. Caution is in order, however — nothing as yet says that Tan and Breen's stripes have anything to do with cortical specification. As before, the open question is which cells are dispersing and which not. Tan and Breen do not say, but warn against expecting there to be a simple split between neurons and glia.

Ironically, just when it seemed safe to throw away models involving boundaries, Fishell and his colleagues (page 636³) find one in the mouse forebrain. By fluorescently marking cells in the VZ in a whole-mount preparation, they follow individual cells over much shorter time periods than those allowed by the retroviral technique. They find that most cells move perhaps 5 or 10 cell diameters in the plane of the VZ during the 8 hours of culture. Some move further, however, perhaps corresponding to the dispersing clones of Walsh and Cepko. The notable observation, however, is a

boundary across which the cells do not move. This is not within the cortex anlage, but between it and the neighbouring lateral ganglionic eminence, the precursor of the striatum. The authors show two cells failing to cross this boundary, although strangely they omit to say how many such cells they observed. There is an apparent conflict here with Walsh and Cepko, who found clones dispersing between cortex and neighbouring regions, but Fishell *et al.* are only considering movement in the VZ, not the later neuronal and glial dispersion.

The implication is that this forebrain boundary is similar to the hindbrain boundaries that have rightly generated such excitement in recent years¹⁰, but it is a little early to say how similar they are. For example, we do not know if the restriction is reciprocal so that striatal cells cannot move into the cortex. Neither do we know if there is an actual barrier, or just a tendency of the

cells not to mix.

So we are faced with both the fact of cell dispersion and the existence of boundaries. Predictably, simplistic models are giving way to more realistic but less tidy hybrids. This is progress, but the cerebral cortex is going to keep us thinking for a while to come. □

Jack Price is at the National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.

1. Walsh, C. & Cepko, C. L. *Nature* **362**, 632–635 (1993).
2. Tan, S.-S. & Breen, S. *Nature* **362**, 638–640 (1993).
3. Fishell, G., Mason, C. A. & Hatten, M. E. *Nature* **362**, 636–638 (1993).
4. Rakic, P. *Science* **241**, 170–176 (1988).
5. O'Leary, D. D. M. *Trends Neurosci.* **12**, 400–406 (1989).
6. Grove, E. A. *Curr. Biol.* **2**, 142–144 (1992).
7. Grove, E. A., Kirkwood, T. B. L. & Price, J. *Neuron* **8**, 217–229 (1992).
8. Walsh, C. & Cepko, C. L. *Science* **255**, 434–440 (1992).
9. Kirkwood, T. B. L., Price, J. & Grove, E. A. *Science* **258**, 317 (1992).
10. Fraser, S., Keynes, R. & Lumsden, A. *Nature* **344**, 431–435 (1990).

ASTROPHYSICS

The tangled web of magnetism

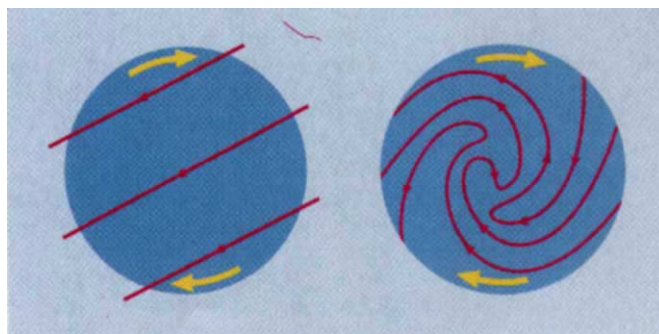
Ellen Zweibel

MAGNETIC fields pervade all galaxies, but their origin is a puzzle. From observations of the synchrotron radiation put out by relativistic electrons spiralling through the fields in galaxies characterized by vigorous star formation, Chi and Wolfendale argue on page 610 of this issue¹ that galactic magnetism has stellar origins.

In the past, analysis of synchrotron radiation from galaxies has been used to infer both the magnetic field strength and the energy density carried by relativistic electrons. At the base of the past analyses has been the assumption that energy is equipartitioned between the particles and the field, an assumption abandoned by Chi and Wolfendale. Equipartition has been an attractive assumption in part because it is easily shown that the total magnetic and particle energy is minimized near equipartition for any given synchrotron luminosity; in the case of the brightest active galaxies, anything that lessens the energy is welcomed by astrophysicists. But furthermore, equipartition seems to hold for our own Galaxy (although one first has to allow for the energy carried by nuclei, 100 times that in electrons, whose synchrotron radiation is negligible).

Chi and Wolfendale have made new

measurements of the active galaxy M82 and of the Magellanic Clouds, the two minor galaxies in orbit around our Galaxy. Their method uses independent γ -ray and X-ray observations to pin



A primordial magnetic field can be wound up by a rotating galaxy into a bisymmetric spiral.

down the particle population independently of the magnetic field strength — the γ -ray flux includes a non-synchrotron component that depends on particle density. They find that equipartition is violated in each of these objects, with magnetic energy density exceeding particle energy density by a full 1–2 orders of magnitude. It is because each galaxy is vigorously forming stars that Chi and Wolfendale suggest a stellar origin for the field.

Three origins for galactic magnetic fields are generally considered. Although they are independent, they

might operate jointly. All three theories build on the fundamental principles that the magnetic field can grow by induction caused by motion in the medium that carries it, and that the magnetic flux through a surface of infinitely conducting material is invariant. (The interstellar gas in galaxies has less than infinite conductivity, of course, but the magnetic flux leakage time is some 10–12 orders of magnitude longer than the age of the Universe.)

The first 'dynamo' theory, the focus of most current work, is based on the idea that galaxies were 'seeded' by very weak, cosmological magnetic fields which were then amplified through some combination of systematic and turbulent gas motions in ionized gas. The role of differential rotation, which is strong in spiral galaxies, is known as the ω effect; turbulence has two roles — induction of field through the so-called α effect, and dissipation of the field through eddy diffusivity (the β effect).

The second possibility is that galaxies are substantially magnetized at the time they form. Amplification of the field occurs as the protogalaxy collapses from the rarified intergalactic gas, and from differential rotation. This wind up of the field into a bisymmetric spiral (see figure) leads to reversals in direction on progressively smaller scales over time, but this problem may be obviated if the charged- and neutral-particle species in the galactic interstellar medium are sufficiently weakly coupled that the magnetic field and ions slip through the neutrals².

Finally, galactic magnetic fields could lack large-scale coherence (observations do not rule this out) and could be produced locally by point sources³ or galactic jets during an epoch of activity⁴. This is what Chi and Wolfendale favour. They argue that galaxies with high star-formation rates have large magnetic fields, whether or not they rotate appreciably, and that magnetic fields in regions of active star formation are disordered and show no mean direction.

This is a bold leap, and unfortunately for the purposes of discriminating between these theories, the correlation of active star formation with strong, disordered fields can be explained at least in part by the havoc wreaked by massive stars on the surrounding interstellar medium. These stars emit powerful ionizing radiation and intense winds over their short lifetimes, blowing bubbles in to the gas and compressing and distorting the magnetic field. When they explode as supernovae, the resulting shock

waves sweep up matter and magnetic fields even more efficiently⁵. Because the intensity of synchrotron radiation is proportional to the square of the magnetic field strength, a field which is spatially intermittent by virtue of having been swept into shells will yield higher synchrotron luminosity than the same amount of magnetic flux in smoothly distributed form, confounding measurements. Furthermore, extensive calculations⁶ have borne out Parker and Ko's suggestion⁷ that the turbulence associated with energy output by massive stars accelerates dynamo activity.

More generally, it appears that the morphology of galactic magnetic fields can tell us little about the fields' origin. When rotation is weak, dynamos of the so-called α^2 type may still operate, and these give rise to fields lacking a dominant azimuthal component, perhaps of the type observed in the Small Magellanic Cloud. Perhaps even more surprisingly, the bisymmetric spiral fields once thought to be conclusive evidence of primordial origin can perhaps be explained by dynamo theories⁸.

How is the question of the origin of galactic magnetic fields to be resolved? Chi and Wolfendale's counterexamples undermining the assumption of energy equipartition are one element; they may also be telling us something about the mechanism of cosmic-ray acceleration. There are also other tantalizing observational clues. For example, whatever large-scale intergalactic field there is cannot be detected at present⁹. Also, Kronberg *et al.*¹⁰ have found magnetism in a distant, dark object, possibly a developing protogalaxy. And the theories are challenged on their own turf as well: primordial field theories must still explain how the Universe came to be magnetized and dynamo theories must answer a host of difficult questions about the operation of turbulent resistivity in the interstellar medium. Perhaps as theorists and observers continue to close in from different directions we will be left with a single viable explanation for galactic magnetic fields. □

Ellen Zweibel is in the Department of Astrophysical, Planetary and Atmospheric Sciences, University of Colorado, Boulder, Colorado 80309, USA.

Satellite maps ozone destroyer

Martyn Chipperfield

OVER the past few years localized measurements of the chlorine monoxide radical in the lower stratosphere from ground-based instruments¹ and aircraft² have revealed that the large springtime loss of ozone over Antarctica is due to catalytic destruction involving this species. Although similar measurements detected comparable levels of ClO in the Arctic winter stratosphere³, the differing meteorology between the two hemispheres has so far prevented the formation of an Arctic ozone hole. Now satellite measurements have revealed the extent of the winter increase in ClO in both polar regions. On page 597 of this issue⁴, Waters *et al.* present the first measurements from the satellite-borne Microwave Limb Sounder, which provide a dramatic four-dimensional picture of the role of ClO in destroying Antarctic ozone and of the potential for similar ozone loss in the north.

More than 80 per cent of the current chlorine loading of the stratosphere (around 3.5 parts per billion by volume, p.p.b.v.) comes from anthropogenic sources such as chlorofluorocarbons. Once these relatively stable chlorine-containing source gases reach the stratosphere they are broken down by solar ultraviolet radiation and by reaction with electronically excited atomic oxygen. According to the known gas-phase chemical reactions, the chlorine thus released should be tied up in the relatively inert reservoir molecules such as HCl and ClONO₂. However, in the winter stratosphere, a strong westerly circulation is set up (the polar vortex) which leads to temperatures cold enough, below about 195 K, for polar stratospheric clouds (PSCs), composed of condensed water and nitric acid, to form.

Reactions can occur on the surface of PSCs to convert these reservoir chlorine species into more reactive forms, a process termed chlorine activation, which ultimately leads to ClO which can efficiently destroy ozone in catalytic cycles (involving Cl₂O₂ and BrO) when sunlight is present⁵. Therefore the requirements for polar ozone depletion are cold temperatures, for PSCs to activate the chlorine, and for this activated air to be exposed to sunlight. These conditions are always met in the Antarctic, but in the Arctic the winters are shorter and warmer, although there is a large amount of year-to-year variability.

Outside the polar regions a downward trend in ozone has been observed at all latitudes except the tropics, with a decrease in the lower stratosphere of 10

per cent per decade⁶. This unexplained downward trend may be caused by transport of ozone-poor air from the polar regions. Or it may be due to *in situ* ClO-catalysed ozone loss either by transport of ClO-rich air from the polar vortex or by chlorine activation on liquid sulphuric-acid aerosols. These aerosols are present throughout the lower stratosphere, and following the eruption of Mt Pinatubo in June 1991 their abundance was significantly increased.

The Microwave Limb Sounder (MLS), used by Waters *et al.*, is one of ten sensors on the NASA Upper Atmosphere Research Satellite (UARS) launched by the space shuttle on 12 September 1991. The satellite was designed to provide an integrated and coordinated study of the chemistry, dynamics and radiation of the Earth's atmosphere. The instruments onboard measure the energy input, winds and composition of the stratosphere, mesosphere and lower thermosphere. In addition to making the first global observations of ClO, MLS measures ozone and water vapour along with pressure and temperature. It is the first space-borne 'limb'-viewing instrument (looking tangentially through the atmosphere) to use the microwave frequency. The planned lifetime of the mission is three years, although problems with the satellite's power supply⁷ might cut it short.

Waters *et al.* present measurements of ClO and ozone from the Antarctic in 1991 and 1992 and from the Arctic in the winter of 1991–92. The Southern Hemisphere data show that the build up of ClO starts early in the season: MLS measured more than 1 p.p.b.v. of ClO in the Antarctic in early June 1992. The observed ozone depletion started in mid-August and later, in the austral spring when the ozone depletion was well underway, the data reveal the striking anticorrelation between high levels of ClO (up to 2.5 p.p.b.v. at 20 km) and low ozone inside the polar vortex. The authors argue that destruction may have started before August, but was hidden by descent of ozone-rich air in the polar vortex.

The Northern Hemisphere measurements show that comparable amounts of ClO were observed in the Northern Hemisphere in early January 1992. The location of the enhanced ClO is consistent with our understanding that chemical reactions on PSCs are responsible for the chlorine activation and the ClO was contained within the polar vortex. Although ozone destruction of 4 per cent was inferred from the ozone data from

1. Chi, X. & Wolfendale, A. W. *Nature* **362**, 610–611 (1993).
2. Kulsrud, R. in *Galactic and Extragalactic Magnetic Fields*, IAU Symp. 140 (eds Beck, R., Kronberg, P. P. & Wielebinski, R.) 527–530 (Reidel, Dordrecht, 1990).
3. Rees, M. J. Q. *J. R. astr. Soc.* **28**, 197–206 (1987).
4. Daly, R. A. & Loeb, A. *Astrophys. J.* **364**, 451 (1990).
5. Ferriers, K. M., Mac Low, M. M. & Zweibel, E. G. *Astrophys. J.* **375**, 239–253 (1991).
6. Ferriere, K. M. *Astrophys. J.* **391**, 188–198 (1992).
7. Parker, E. N. & Ko, C. M. *Astrophys. J.* **341**, 828 (1989).
8. Krause, M. in *Galactic and Extragalactic Magnetic Fields*, IAU Symp. 140 (eds Beck, R., Kronberg, P. P. & Wielebinski, R.) 187–196 (Reidel, Dordrecht, 1990).
9. Vallee, J. P. *Astrophys. J.* **360**, 1–6 (1990).
10. Kronberg, P. P. *et al. Astrophys. J.* **387**, 528 (1992).

1–13 January 1992, temperatures in the Arctic stratosphere rose above the PSC threshold later in the month, at which point the ClO levels decayed, by conversion to ClONO₂ (ref. 8), and large ozone loss was avoided.

An interesting feature of the MLS data is the magnitude of the measured ClO. In both polar regions, MLS measured ClO mixing ratios of up to 2.5 p.p.b.v. at around 20 km altitude. This is approaching the limit of the expected total amount of chlorine in the atmosphere. Indeed, as some chlorine is expected to be in the form of the dimer of ClO (which is in equilibrium with ClO and is not measured by MLS) these measurements may be hard to reconcile with the known photochemical data of the ClO/Cl₂O₂ system.

Although Arctic ozone loss was limited in the winter of 1991–92, the natural variability in the length of the Northern Hemisphere winter could prolong the occurrence of PSCs into late February or even March. In the winter just passed, PSCs did persist into late February giving greater opportunity for ozone loss. Waters *et al.* comment briefly on this period and note that they measured ClO higher than 1 p.p.b.v. until 4 March 1993. In late February, they observed ozone inside the polar vortex to decrease by 0.7 per cent per day. On the basis of the colder winter and prolonged period of high ClO, substantial ozone loss should be expected in the Arctic spring of 1993. Indeed, on the basis of satellite- and ground-based observations, the World Meteorological Organisation reported that over much of the Northern Hemisphere throughout February ozone levels were up to 20 per cent below normal. A quantitative study will be needed to see if such polar ozone loss can explain all of the ozone decrease seen by MLS or whether other processes, which may be causing the gradual ozone loss at mid-latitudes, are contributing. The global data provided by MLS along with the other instruments on UARS (to appear in a special issue of *Geophysical Research Letters*) will be an invaluable aid in trying to understand and quantify the evolution of stratospheric ozone. □

Martyn Chipperfield is at the Centre for Atmospheric Science, Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK.

Instant patterns in thin films

Daniel K. Schwartz

WHAT do magnets and soap films have in common? The answer is spontaneous symmetry breaking. Physicists ascribe permanent magnetization of iron to a broken time-reversal symmetry. And on page 614 of this issue¹, Eckhardt *et al.* show that a thin film of soap-like organic molecules can separate into regions of different handedness, breaking the inherent mirror symmetry of the system.

into right- and left-handed regions, thus breaking the natural mirror symmetry. This effect is related to ferroelectricity at interfaces.

Eckhardt *et al.* report the first direct observation of this phenomenon on molecular length scales. They have imaged the molecular lattice of an organic monolayer using atomic-force microscopy and find that, under certain condi-

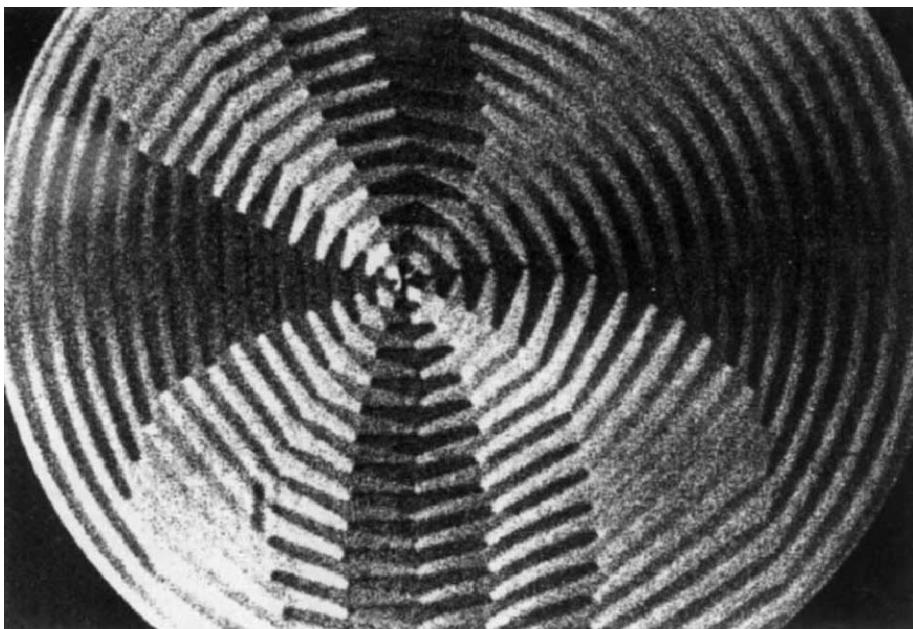


FIG. 1 A pattern on a thin, freely suspended liquid-crystal film observed through crossed-polarizers. The different shades of the stripes are due to the azimuthal tilt direction of the rod-like molecules. The soliton (sharp-edged) stripes predicted by theory⁸ are arranged into a twelve-armed star. (From ref. 5.)

Although nature has a preference for symmetry, many physical systems adopt ground states that lack the full symmetry that they might, in principle, assume. The most familiar example of this spontaneous symmetry breaking is ferromagnetism. In the absence of an external magnetic field, the unpaired spin on each atom in a lump of iron can be either up or down with equal probability. But at low temperatures, spins within a macroscopic domain align to produce a finite magnetization, thus breaking the inherent up/down symmetry. This occurs even without the presence of an ordering magnetic field, hence the word spontaneous.

Another type of symmetry often found in physical systems is chiral, or mirror, symmetry. Pasteur's work on optical activity in sugars in the 1860s demonstrated that chiral effects in liquids are related to chirality at the molecular level. However, recent studies show that two-dimensional liquids composed of achiral units can spontaneously separate

tions, a monolayer composed of equal numbers of right- and left-handed molecules (a racemic mixture) can separate into chiral regions.

Other studies have revealed remarkable patterns on scales 10⁴ times larger (10 μm or more) in quasi-two-dimensional systems such as thin freely suspended liquid-crystal films and Langmuir monolayers (monolayers of amphiphilic molecules at the air/water interface). Striking chiral patterns such as spirals or pinwheels are found in systems composed of chiral molecules². Clearly, the chiral nature of such patterns is the macroscopic manifestation of the underlying molecular chirality. And recently there have been reports of chiral domain shapes in racemic³ and even achiral⁴ systems. Striking patterns such as stripes, honeycombs and striped stars (Figs 1 and 2) have also been observed and attributed to spontaneous chiral separation in systems of achiral molecules^{5,6}. The exciting possibility is that Eckhardt *et al.* have observed the

1. de Zafra, R. L. *et al.* *Nature* **328**, 408–411 (1987).
2. Anderson, J. G. *et al.* *J. geophys. Res.* **94**, 11465–11479 (1989).
3. Brune, W. H. *et al.* *Geophys. Res. Lett.* **17**, 505–508 (1990).
4. Waters, J. W. *et al.* *Nature* **362**, 597–602 (1993).
5. Solomon, S. *Nature* **347**, 347–354 (1990).
6. Stolarski, R. *et al.* *Science* **256**, 342–349 (1992).
7. *Science* **259**, 883 (1993).
8. Oelhaf, H. *et al.* *Geophys. Res. Lett.* (EASOE special issue, in the press).

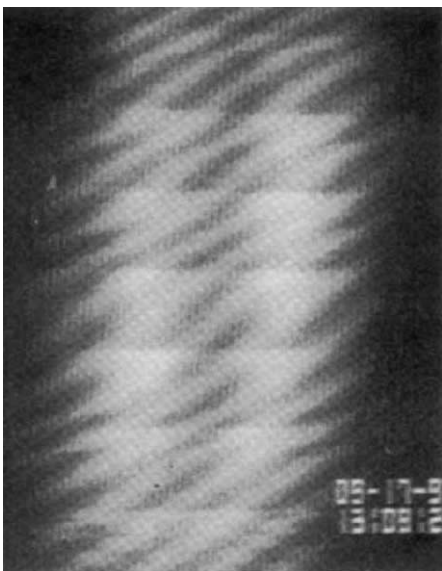


FIG. 2 A stripe pattern observed on a Langmuir monolayer of a fatty acid by polarized fluorescence microscopy. The slow variation in shade across each stripe is due to a continuous change in the molecular azimuthal tilt direction. Such continuous variation suggests that the monolayer is in a hexatic phase, as it should not occur in a true two-dimensional crystal. (Courtesy of C. M. Knobler.)

molecular lattice of tiny parts of such a pattern.

There are several mechanisms by which chirality can be introduced into a system of inherently achiral units. The achiral unit can separate into two chiral constituents. This may be what is seen by Eckhardt and colleagues: the two molecular enantiomers normally combine to make an achiral dimer, but can be induced to separate into distinct phases. Alternatively, an achiral molecule can choose to pack in a chiral lattice. Such a packing has been observed in a thin Langmuir-Blodgett film of an achiral fatty-acid salt⁷.

Also, rod-shaped molecules packed hexagonally (a physicist's conception of a liquid crystal) can form a tilted phase in which the tilt direction is neither directly towards a nearest-neighbour nor exactly between nearest-neighbours, thus breaking the mirror symmetry. This is the structure of the liquid-crystal smectic L phase which has long-range bond-orientational order but only short-range translational order (a hexatic phase).

A recent theoretical paper by Selinger *et al.*⁸ describes a phase diagram for the patterns formed by a chiral hexatic system that predicts striped patterns as well as two-dimensional modulation giving, for example, a square pattern. Their theory predicts two types of stripes, one in which the order parameter varies continuously and one with sharp domain walls. Changes in temperature can alter the stripe widths as well as induce

transitions between patterns.

There is striking qualitative agreement between the predictions of Selinger *et al.* and the observations by Maclennan and Seul of patterns formed on freely suspended liquid-crystal films⁵. These authors observed the evolution from sharp stripes to a two-dimensional modulation (a honeycomb) with increasing temperature under conditions at which the surface layers of the film are in a hexatic phase. They also found intriguing and regular cusp defects in the stripe patterns and the twelve-armed star pattern already mentioned (Fig. 1), both of which will serve as a further challenge for theory.

The situation is not so clear for the patterns formed on Langmuir monolayers. The stripes observed at the air/water interface are neither a sine wave nor a square wave (sawtooth might be a better description of the pattern in Fig. 2). Perhaps more problematically, as the microscopic structure of the phase is not known with certainty, it is unclear if the phase is chiral or even if it is hexatic. An additional complication is that Eckhardt *et al.* observe their monolayer after transferring it onto a solid support. These factors make it difficult to apply Selinger's theory directly.

At the most basic level, the connection between chiral symmetry breaking and pattern formation is of interest because it is another example of the ways in which structure and symmetry at the molecular level make themselves felt macroscopically. Perhaps the most dramatic and amazing example of this is crystal faceting. There may be many technological uses for regular patterns — in displays, gratings and electronic devices, for example. If we can understand the relationship between molecular architecture and the patterns formed it will be the first step towards engineering molecules to form the types of patterns we need. The huge range of length scales (1 Å to 100 μm) that can be observed with the atomic-force microscope and other scanning probe microscopes make it possible to explore this question directly. □

Daniel K. Schwartz is in the Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90024, USA.

1. Eckhardt, C. J. *et al. Nature* **362**, 614–616 (1993).
2. Gaub, H. E., Moy, V. T. & McConnell, H. M. *J. phys. Chem.* **90**, 1721–1725 (1986).
3. Stine, K. J., Uang, J. Y.-J. & Dingman, S. D. *Langmuir* (submitted).
4. Qiu, X., Ruiz-Garcia, J., Stine, K. J. & Knobler, C. M. *Phys. Rev. Lett.* **67**, 703–706 (1991).
5. Maclennan, J. & Seul, M. *Phys. Rev. Lett.* **69**, 2082–2085 (1992).
6. Qiu, X., Ruiz-Garcia, J. & Knobler, C. M. *Mat. Res. Soc. Symp. Proc.* **237**, 263–270 (1992).
7. Schwartz, D. K., Viswanathan, R. & Zasadzinski, J. A. N. *Phys. Rev. Lett.* **70**, 1267–1270 (1993).
8. Selinger, J. V., Wang, Z.-G., Bruinsma, R. F. & Knobler, C. M. *Phys. Rev. Lett.* **70**, 1139–1142 (1993).

Hereditary bias

LAST week Daedalus suggested that a woman's two ovaries compete each month for the chance to ovulate. He now explains why. Each ovary, he reckons, has its own genetic strategy. It will split the 24 human chromosome pairs with a certain bias, preferentially retaining some chromosomes in the ovum while discarding others. Meanwhile the other ovary, like a rival college, is turning out graduates with a different taste and style. Each month, the ovaries compete for the honour of putting up one of their own graduates for a potential pregnancy.

And this, says Daedalus, is the point of having two ovaries. Like the left and right sides of the brain, they have different strengths. Their owner instinctively judges her social circumstances and potential mates, and unconsciously shifts her hormonal balance to release ova more often from the favoured ovary. In ordered circumstances, for example, the left ovary might be preferred; it would release ova with a logical and systematic endowment. In more chaotic times, the right ovary's graduates would promote the genetic bases of opportunism and creative flair.

Each ovum seeks to implement the genetic strategy of its home ovary. Launched into a mass of eager sperms, it will scan its suitors critically, looking for the best genetic match. It may not find it. It is the extreme choosiness of ova, says Daedalus, that makes pregnancy so uncertain.

The male response (as ever) is to play the field. 24 chromosome pairs can be divided up to generate 2²⁴ different possible sperms: about 17 million, a good order-of-magnitude estimate of an ejaculatory content. And this is why the male makes so many sperms. He is covering all genetic permutations, so that even the choosiest ovum should find its ideal partner among them. Yet even here there is room for strategy. Left and right testicles may also bias their output in the light of the prevailing conditions, to maximize the competitive fitness of the resulting offspring.

This ability to bias our children's character to suit the immediate circumstances, gives us a mighty advantage over strictly mendelian creatures. Indeed, says Daedalus, this must be the secret of the speed and vigour of human evolution, and the wonderful adaptability of our species. He is now seeking evidence of sperm preferences among ova *in vitro*. An ideally choosy ovum should reject the first (1/e)th of the candidates, while noting their characters. Having thus sized up the field, it should accept the next sperm that is superior to any of the previous sample.

David Jones

Should a lion change its spots?

SIR — Fluctuating asymmetries can provide important insights into patterns of mate choice^{1,2} because asymmetric individuals often show low levels of overall genetic health³. But the complex relationship between fluctuating asymmetry and lifespan in African lions suggests that this trend may not be universal.

Lions, like other cats, have conspicuous vibrissae spots on either side of their muzzle. Each spot is a small furless area that surrounds a single 'whisker'. Parallel rows of vibrissae spots are located on each side of each lion's face. Above the topmost row, most lions possess a few additional identification spots, which we recorded systematically when an individual was first observed (usually as a

small cub). The precise patterns of these spots are highly variable and vibrissae spots are used for individual recognition in most field studies of lions⁴. We have now examined fluctuating asymmetry in the whisker-spot patterns of 920 lions from our long-term studies⁵⁻⁷.

A lion may differ in the number of identification spots on each side of its face or in the location of spots above the topmost 'reference' row, but these patterns do not differ consistently between left and right. Nor is there any correlation between the asymmetry of a mother's and her cubs' spots. Of 793 lions whose spot patterns can be scored for the number and location on both sides of the face, 759 (96%) are unique.

We tested whether any measure of spot-pattern asymmetry is correlated with two factors known to affect fluctuating asymmetry³: the extent of close inbreeding and the intensity of nutritional stress during development. The Ngorongoro crater lions are inbred and show significantly reduced levels of genetic variability^{5,6}, whereas the lions on the Serengeti plains are subject to far greater variation in food supply than lions in either the Serengeti woodlands or the crater⁷. However, the extent of asymmetry in identification spots does not differ across habitats, nor is spot asymmetry correlated with isozyme heterozygosity.

We tested whether asymmetry in the number or horizontal location of identification spots is correlated with specific components of lifetime reproductive success. We restricted our analysis to individuals that died during the course of our study but lived for a minimum of 5 years, by which age both sexes have generally become established breeders⁷.

We found no relationship between fitness and asymmetry in the number of spots on each side of the face. However, lifespan in both sexes is significantly correlated with the difference in horizontal location of their identification spots (see figure). Even more surprising, the correlation runs in the opposite direction in each sex. Consistent with other studies of fluctuating asymmetry, asymmetric males die at a younger age than symmetrical males; but asymmetric females live longer than symmetrical females. These results are statistically significant both by simple regression and in multiple-regression models that include the main factors known to influence longevity in the two sexes⁷.

Female mammals that are heterozygous for genes on the X chromosome often have mosaic skin patterns in either coat colour (for example, calico cats) or in the distribution of sweat glands

(anhidrotic ectodermal dysplasia) due to X-chromosome inactivation (a phenomenon generally, and in this case rather amusingly, known as lyonization). However, our results do not result from lyonization: the degree of spot asymmetry is the same in the two sexes. Thus asymmetric spot patterns in lions are probably generated by a similar mechanism in both sexes.

It is difficult to accept that these correlations reflect direct causality between spot asymmetry and survival. It is more likely that spot asymmetries reflect some underlying genetic or physiological condition expressed during early development that is subject to opposing selective forces in the two sexes. Whatever the underlying mechanism, male and female lions would do better to show divergent preferences in the pattern of their partners' spots.

Craig Packer

Anne E. Pusey

*Department of Ecology, Evolution and Behavior,
University of Minnesota,
St Paul, Minnesota 55108, USA*

1. Møller, A. P. *Nature* **357**, 238–240 (1992).
2. Thornhill, R. *Behav. Ecol.* **3**, 277–283 (1992).
3. Palmer, A. R. & Strobeck, C. A. *Rev. Ecol. Syst.* **17**, 391–421 (1986).
4. Pennycuik, C. & Rudnai, J. J. *Zool.* **160**, 497 (1970).
5. Packer, C. *et al. Conserv. Biol.* **5**, 219–230 (1991).
6. Wildt, D. E. *et al. Nature* **329**, 328–331 (1987).
7. Packer, C. *et al. in Reproductive Success* (ed. Clutton-Brock, T. H.) 363–383 (Univ. Chicago Press, 1988).

More word play

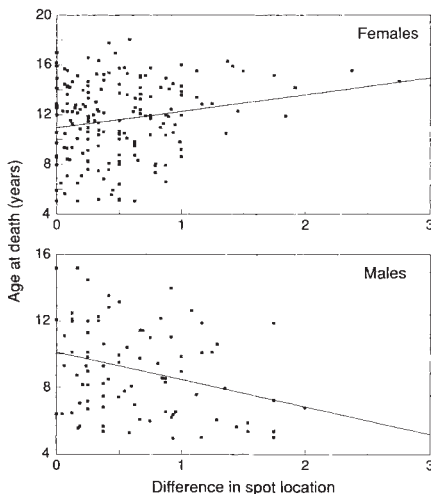
SIR — In the light of recent searches for words from languages in protein sequences transcribed by the single-letter amino acid code¹, and in the spirit of Harvey's suggestion², we report our findings from the circumsporozoite protein of the malaria parasite *Plasmodium falciparum*. This protein is expressed at a high level on the surface of the infective sporozoite stage of the parasite and elicits a strong but relatively ineffective T-cell-independent immune response on the part of the host. Thus it has been argued that the function of the circumsporozoite protein is precisely to elicit such an immune attack³. Appropriately, we find in the carboxy-terminal region of the molecule the English phrase 'KICKME'. This is a conserved sequence motif, found in all alleles of this protein sequenced to date.

Austin L. Hughes

Marlaine K. Hughes

*Department of Biology and Institute of
Molecular Evolutionary Genetics,
The Pennsylvania State University,
University Park, Pennsylvania 16802, USA*

1. Gonnet, G. H. & Benner, S. A. *Nature* **361**, 121 (1993).
2. Harvey, S. C. *Nature* **362**, 300 (1993).
3. Kemp, D. J., Koppel, R. L. & Andrews, R. F. A. *Rev. Microbiol.* **41**, 181 (1987).



Regression of lifespan against horizontal identification spot asymmetry for females (top) and males (bottom). The horizontal location of each spot is scored according to its position above the topmost row of vibrissae spots, or 'reference' row, on that side of the lion's face. Reference-row spots are numbered sequentially from the spot closest to the tip of the lion's nose. If a lion has two spots on the left side of its face directly above the first and second reference-row spots, its position score for that side would be scored as $(1+2)/2=1.5$. If it has one spot directly above the third reference-row spot on its right side, the horizontal difference in spot location would be $3-1.5=1.5$ and the difference in the number of spots would be $2-1=1$. Longevity is significantly related to horizontal asymmetry in both sexes: for females, the least-squares linear regression of age at death against horizontal spot differences: $r^2=0.0378$, $n=176$, Student's $t=2.61$, $P<0.01$. Female lifespan varies across habitats, but the effect of horizontal spot difference remains significant when habitat is included in a multiple-regression model: $t=2.78$, $P<0.01$. For males: $r^2=0.0844$, $n=87$, Student's $t=-2.80$, $P<0.01$. Males in larger coalitions live longer, but horizontal spot asymmetry remains significant in a multivariate model: $t=-2.84$, $P<0.01$.

Last interglacial in Devils Hole

SIR — Broecker in *News and Views*¹ confessed himself confused by the apparent contradiction between the chronologies for the last interglacial presented by the Devils Hole (Nevada) calcite vein² and that generally accepted by adherents to the Milankovitch theory³. I suggest that the discrepancy arises from an unforeseen consequence of dating this unusual deposit by measuring the growth of ^{230}Th from its parent ^{234}U .

The oxygen-isotope record from core DH11 (ref. 4), obtained from a calcite crust in the Devils Hole, provides an apparently amazing confirmation of the global validity of the continuous marine $\delta^{18}\text{O}$ records as a description of the first order pattern of the ice ages. Almost every detail in the SPECMAP stacked $\delta^{18}\text{O}$ record³ can be recognized in this precisely dated core which is only 36 cm long (Fig. 3 of ref. 4). The one apparent discrepancy that leads these authors² eloquently to disagree with the authors of the SPECMAP paper³ is a seemingly crucial disagreement in the timing of the onset of the last interglacial.

The Devils Hole is a water-filled fault fissure about 1.2 m wide lined with a thick layer of calcite; it is this layer of calcite that was sampled. The calcite and associated uranium has been deposited from the water. Ages are estimated from the ingrowth of ^{230}Th in apparently technically ideal conditions². The average accumulation rate is about 0.6 mm per kyr, although over the critical interval of interest 5.5 mm accumulated between 149.8 and 131.9 kyr, a rate of about 0.3 mm per kyr. If we envisage the water from half the width of the fissure contributing to the calcite on each side, then out of the 0.6 m thick layer of water, calcite was deposited at a rate of 0.8×10^{-4} g cm⁻² of surface per year. The calcite contains about 0.5 p.p.m. uranium (Table 1 of ref. 5) so that uranium is deposited at a mean rate of 0.4×10^{-10} g cm⁻² of surface per year.

An important difference between the Devils Hole situation and the cave speleothems that have usually been dated by this method (reviewed in ref. 6) is that the vein is continuously in water. This difference is important because one must consider the fate of the ^{230}Th that

is produced by decay of dissolved ^{234}U in the water. A layer of water 0.6 m thick and containing 3 p.p.b. uranium (ref. 7 and I. J. Winograd, personal communication) represents 0.18×10^{-6} g of uranium per cm² of the fissure face, and thus will generate in one year a quantity of ^{230}Th about 4.5×10^5 times the amount generated by the decay of the uranium in one year's accumulation on the vein. Because thorium is very insoluble, this ^{230}Th will immediately attach to the surfaces of the cavern, so that even though the ^{232}Th concentration of the calcite is negligible, zero-age calcite could have an apparent age of the order of 5×10^3 yr if age is estimated on the assumption that freshly deposited calcite contains no ^{230}Th . Although this ^{230}Th excess will decay with time, its contribution to the apparent age remains a constant number of years. Moreover, the initial age error would vary inversely with the growth rate; it would be difficult to interpret the apparent 60 kyr age obtained for the outer part of the calcite where calcite accumulation must have slowed towards zero with an ever increasing "initial age", although the excellent $^{234}\text{U}/^{238}\text{U}$ data provide a valuable constraint.

I conclude that rather than being a thorn in our flesh, the record generated by Winograd and colleagues⁴ is both the most impressive record of the ice ages from the American continent, and a remarkable demonstration of the ubiquity of the characteristic record that has been so widely documented in the sediments of the deep ocean. The general validity of the timescale currently in use is strongly supported by the impressive ^{234}U chronology², but despite the technical excellence of the ^{230}Th measurements, I suggest that the potentially large contribution from unsupported ^{230}Th precludes their use to refine our best estimate for the age of the last interglacial.

N. J. Shackleton

Godwin Laboratory for Quaternary Research,
University of Cambridge,
Cambridge CB2 3RS, UK

LUDWIG ET AL. REPLY — Our disagreement with SPECMAP and with the Milankovitch hypothesis stems not only from the "timing of the onset of the last interglacial", as claimed by Shackleton, but also from the following features of the Devils Hole record: the 20-kyr duration of the interglacial climates; an increasing duration of the quasi-100-kyr cycle as we come forward in time; and the occurrence of a well-developed glacial cycle at 350–450 kyr when the insolation theory indicates none should occur¹.

With regard to the ^{230}Th excess issue,

properly raised by Shackleton, we note that: (1) groundwater flow in Devils Hole is laminar, whereas turbulent flow would be required to efficiently transfer water-generated ^{230}Th to the cavern walls; (2) particulate scavenging of ^{230}Th , followed by gravitational settling, is strongly suggested by the contrast between the textures of footwall vein calcite compared to hanging wall vein calcite. (Footwall calcite in Devils Hole is coloured and finely banded, whereas hanging wall calcite — such as samples DH-11 and DH-2 — is white-to-yellowish white and rarely banded.); (3) the necessity of allowing for groundwater travel time means that the age of the onset of the last interglacial recorded in DH-11 (140 ± 3 kyr) must itself be too young by several thousand years; and (4) most conclusively, since receipt of an earlier draft of Shackleton's correspondence, we have measured the ^{230}Th - ^{234}U - ^{238}U of the outermost surface (to a depth of 40–50 μm) of vein samples DH-2 and DH-11 and found only a negligible fraction (<1%) of the excess ^{230}Th predicted by Shackleton. These analyses⁸ show that the effect of water-generated excess ^{230}Th , accumulated on the surface of these veins during the past 60 kyr, corresponds to an age-bias of less than 20 years over the entire history of vein calcite accumulation for DH-11 and DH-2.

Shackleton reasonably asks whether ^{230}Th in the Devils Hole water column could have significantly affected the accuracy of the ages we reported^{2,4}. However, neither the character of the aqueous environment, the petrology of the calcite, nor direct measurements of excess ^{230}Th on the surface of the veins supports the 5 kyr age-bias suggested by Shackleton.

The Devils Hole dates^{2,4} therefore remain a challenge to the Milankovitch hypothesis.

K. R. Ludwig

K. R. Simmons

I. J. Winograd

B. J. Szabo

J. M. Landwehr

A. C. Riggs

US Geological Survey,
Denver Federal Center,
Lakewood,
Colorado 80225, USA

1. Broecker, W. S. *Nature* **359**, 779–780 (1992).
2. Ludwig, K. R. et al. *Science* **258**, 284–287 (1992).
3. Imbrie, J. et al. in *Milankovitch and Climate* (eds Berger, A. et al.) part 1, 269–305 (Reidel, Dordrecht, 1984).
4. Winograd, I. J. et al. *Science* **258**, 255–260 (1992).
5. Winograd, I. J. et al. *Science* **242**, 1275–1280 (1988).
6. Gascoyne, M. & Schwarcz, H. P. in *Uranium Series Disequilibrium: Applications to Environmental Problems* (eds Ivanovich, M. & Harmon, R. S.) 268–300 (Clarendon, Oxford, 1982).
7. Osmond, J. K. & Cowart, J. B. in *Uranium Series Disequilibrium: Applications to Environmental Problems* (eds Ivanovich, M. & Harmon, R. S.) 202 (Clarendon, Oxford, 1982).
8. Ludwig, K. R. et al. *Science* **259**, 1626 (1993).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

Stratospheric ClO and ozone from the Microwave Limb Sounder on the Upper Atmosphere Research Satellite

J. W. Waters, L. Froidevaux, W. G. Read, G. L. Manney, L. S. Elson, D. A. Flower, R. F. Jarnot & R. S. Harwood*

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

* Department of Meteorology, Edinburgh University, Edinburgh EH9 3JZ, UK

Concentrations of atmospheric ozone and of ClO (the predominant form of reactive chlorine responsible for stratospheric ozone depletion) are reported for both the Arctic and Antarctic winters of the past 18 months. Chlorine in the lower stratosphere was almost completely converted to chemically reactive forms in both the northern and southern polar winter vortices. This occurred in the south long before the development of the Antarctic ozone hole, suggesting that ozone loss can be masked by influx of ozone-rich air.

THE conversion of anthropogenic chlorine in the stratosphere to chemically reactive forms that deplete ozone¹ has been of increased concern since the discovery² of large springtime ozone loss over Antarctica, and findings^{3,4} that the loss is due to chlorine chemistry. Heterogeneous chemistry on polar stratospheric clouds (PSCs) followed by the action of sunlight converts stratospheric chlorine from fairly inert to very reactive forms, ClO being dominant⁵⁻⁷. Enhanced ClO has been observed to precede the Antarctic ozone loss, with correlation between high ClO and low ozone developing in mid-September⁴. Enhanced ClO in the Arctic has also been observed⁸ and associated ozone loss detected⁹. Downward ozone trends at all latitudes outside the tropics have been detected in all seasons¹⁰.

The Microwave Limb Sounder (MLS), launched on 12 September 1991 on the Upper Atmosphere Research Satellite (UARS)¹¹, is obtaining global measurements of ClO. Ozone, H₂O, temperature and pressure are simultaneously measured. MLS is the first use of microwave limb sounding¹² from a satellite, and here we report the main results for ClO and ozone obtained so far. These results show that in the north, as in the south, the existing chlorine in the lower stratosphere can be virtually completely converted to reactive forms throughout the polar winter vortex. The period of enhanced ClO in the 1991-92 northern vortex, and consequent destruction of ozone, was limited by the lower-stratosphere warming above the PSC threshold in late January, before the longer duration of high-latitude sunlight needed to sustain the ozone destruction cycle. A vortex-integrated ozone loss of ~4% in a lower-stratospheric layer during 1-13 January is calculated from the observed ClO. Measurements of ClO in the south during early and mid-winter have now been made by MLS, and show that ClO within the southern winter polar vortex was enhanced by 1 June 1992, long before the 'springtime' Antarctic ozone hole. A decrease of ozone coinciding with the largest ClO abundances was observed starting in mid-August, and MLS maps have now shown the full extent of coincidence between enhanced ClO and depleted ozone in the Antarctic in late September. Measurements in both hemispheres indicate the difficulty of directly observing ozone loss associated with enhanced ClO in early and mid-winter when an influx of ozone-rich air can be simultaneous with the conversion of chlorine to reactive forms. The presence of enhanced ClO coincident with increasing ozone within the polar vortices implies more overall chemical loss than would be suggested from just the late winter and early spring decrease in Antarctic ozone. This raises the question of the extent to which the implied

additional loss is distributed as a longer-term (and perhaps more widespread) ozone reduction. Early 1993 results for the north, when stratospheric temperatures were below the PSC threshold until at least late February, show enhanced ClO for much longer than in 1992, and lower values of late winter ozone abundances throughout most of the Northern Hemisphere. The overall results obtained during the first 18 months of MLS operation indicate that chlorine depletion of stratospheric ozone is of greater concern than previously thought.

The UARS Microwave Limb Sounder

The instrument¹³ has three radiometers simultaneously measuring millimetre-wavelength thermal emission in spectral bands near 205 GHz (for ClO, ozone, SO₂ and upper tropospheric H₂O), 183 GHz (H₂O and ozone) and 63 GHz (temperature and pressure). Measurement coverage is from 80° on one side of the Equator to 34° on the other. There are 15 orbits per day, and the orbit plane precesses slowly with respect to the Sun-Earth direction, so that local solar time of MLS measurements at a given latitude (on either the 'day' or 'night' side of the orbit) varies by only 20 minutes during a 24-hour period. UARS performs a 180° 'yaw manoeuvre' when the orbit precesses through 180° (10 times per year), so MLS high-latitude coverage switches between north and south every ~36 days.

Measurements are made continuously day and night, and are not degraded by stratospheric clouds or aerosols. The composition measurements are fairly insensitive to variations or uncertainties in atmospheric temperature. Overall calibration accuracy¹⁴ is ~3%, and radiometric calibration is done on each limb scan by observing an ambient black-body target and cold space. Figure 1 shows examples of measured ClO spectra. Spectra from the upper stratosphere (Fig. 1a, b) show excellent agreement between balloon and UARS MLS measurements made at the same latitude and season. Spectra from the lower stratosphere (Fig. 1c, d) indicate comparable ClO abundances over Antarctica on 21 September 1991 and over Moscow on 11 January 1992.

Vertical profiles of atmospheric parameters are retrieved using optimal sequential estimation¹⁵ on spectra from each 65-s limb scan, with no retrieval 'memory' between different scans. Figure 2 shows averages of retrieved ClO profiles for three periods in the 1991-92 Northern Hemisphere winter. Large abundances of lower-stratospheric ClO are evident at high latitudes during January. These enhanced ClO abundances can only be explained by heterogeneous chemistry, and can cause significant ozone

depletion⁷. The 30–40° N lower-stratospheric ClO abundances shown in Fig. 2 agree better with models^{16–18} that include sulphate aerosol chemistry¹⁹ than with those containing only gas-phase chemistry, as has also been found²⁰ by aircraft measurements at these latitudes. Although the ClO increase expected from sulphate aerosol chemistry is too small, ~ 0.1 – 0.2 p.p.b.v. (parts per 10^9), to be seen in MLS individual profile measurements, zonal averages of the data over a range of latitudes and seasons will provide valuable information for improving our understanding of this chemistry and its global implications. The upper-stratospheric ClO peak abundances around 2–5 hPa agree well with recent balloon measurements at submillimetre wavelengths²¹.

ClO in the 1991–92 northern winter

MLS daily maps showed enhanced localized ClO abundances greater than ~ 1 p.p.b.v. in the north starting around 14 December 1991. These occurred on the northern edge of daylight measurements near 30° E, 60° N (northern Russia). National Meteorological Center (NMC) data show temperatures below 195 K (roughly the threshold for type-I nitric acid trihydrate PSCs^{22,23}) north of this region for several days previously, and indicate southward advection of air into the area where higher concentrations of ClO were observed. By early January 1992, the enhanced ClO had become comparable, both in magnitude and areal extent, to that measured by MLS late September 1991 in the Antarctic ozone hole. Figure 3 shows lower-stratospheric ClO and ozone extracted from the retrieved profiles for 11 January 1992, and mapped on the 465 K potential temperature (θ) surface. Adiabatic flow occurs on constant θ , and atmospheric pressure at $\theta = 465$ K on this day varied between ~ 45 – 65 hPa in the Arctic vortex (see below) to ~ 75 – 85 hPa in the Aleutian high. The maps were generated using Fourier transform techniques to help to separate temporal and spatial variations^{24,25}. Accuracies of ~ 0.4 p.p.b.v. for ClO and ~ 0.3 p.p.m.v. (parts per 10^6) for ozone are expected for the maps shown here. MLS data from only the south-going 'day' side of the orbit, where measurement local times varied from 10:30 a.m. at 70° N, to 12:30 p.m. at 40° N, were used (the 'night' side of the orbit shows much smaller ClO abundances, as expected because there is no photolysis of the ClOOCl formed when two ClO molecules combine). United States and European Arctic measurement campaigns have also found high ClO in the 1991–92 northern winter (personal communications from campaign members).

Figure 3 also shows 11 January 1992 maps of temperature, Rossby–Ertel potential vorticity (Q) and winds produced from NMC data. Intersections of Q and θ mark material lines for adiabatic frictionless fluid flow and provide useful diagnoses of the material field evolution²⁶. Strong Q gradients bound the vortex that dominates high-latitude winter stratospheric circulation²⁶, and maps of Q on isentropic (constant θ) surfaces, as in Fig. 3, help to identify the vortex location and structure. Geopotential height and temperature fields produced operationally by NMC are routinely used to calculate Q on isentropic surfaces for daily comparison with MLS maps. The horizontal wind field is estimated using a primitive equation balance, neglecting time tendency and vertical advection terms²⁷, with the calculation done on the NMC polar stereographic grid. The winds and temperatures are interpolated by a cubic spline from standard NMC pressure levels to the isentropic surfaces on which Q is calculated. Q was also calculated using data from the UK Meteorological Office assimilation system²⁸; this agreed in all general characteristics with that calculated from NMC data for this period.

The enhanced ClO is generally bounded on the north by the edge of daylight, and on the south by $Q \approx 2.5 \times 10^{-5} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$. Note the displacement of the vortex and enhanced ClO away from the pole towards Russia, the arm extending over Britain, and longitudinal variations inside the vortex. (The patches of enhanced ClO seen at mid-latitudes are under investigation; instrument noise can produce occasional spurious values of their magnitude.) Minimum temperatures occur over northern Scandinavia, where ClO increases and coincident measurements by the UARS CLAES and ISAMS infrared instruments show the greatest density of PSCs. The sharp ClO increase from night to day and the further increase in the PSC region over Scandinavia seem generally consistent with current understanding⁷ of the diurnal chemistry and PSC activation of chlorine. The ClO decrease within the vortex away from the PSC region is correlated with northward excursion of Q contours, indicating air movement to lower solar illumination which would cause less ClOOCl photolysis and ClO. ClO decrease could also be caused by chemical conversion of reactive chlorine back to reservoir forms. NO_2 rapidly reacts with ClO to form ClONO_2 (time constant of ~ 10 minutes at 50 hPa and 195 K for 1 p.p.b.v. each of NO_2 and ClO), and can be produced by photolysis²⁹ of HNO_3 from evaporating PSC. Q contours coincident with high ClO enter the polar night at ~ 210 K temperature near ~ 120 – 150° E, where 0.7 p.p.b.v. ClO is observed

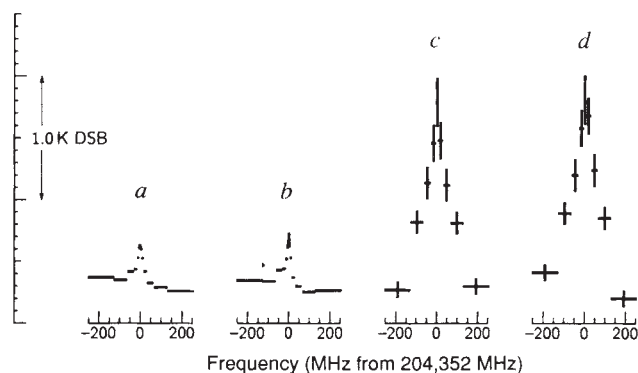


FIG. 1 ClO thermal emission spectra. *a*, Measured by the balloon MLS on 26 September 1989 for 34° N, 33–34 km tangent height and 20 minutes integration time (R. A. Stachnik, personal communication); *b*, average of measurements by UARS MLS on 21 September 1991 at 20–34° N latitudes and 30–40 km tangent height with 3 minutes integration; *c*, measured 21 September 1991 with 1.8-s integration over the Antarctic Weddell Sea at 20 km tangent height—the spectral line is broader than in *a* and *b* because of larger pressure broadening at the lower heights; *d*, measured 11 January 1992 with 1.8-s integration over Moscow at 20 km tangent height. The vertical axis is the emission brightness temperature in 'double-sideband' Kelvin units. Vertical extents of the crosses show $\pm 1\sigma$ measurement noise; horizontal extents show spectral resolution. Some central spectral channels have been averaged together in *c* and *d*. UARS MLS measures spectral points simultaneously in 90 spectral channels, of which only the 15 channels covering ClO are shown here. Its field of view (FOV) is step-scanned through the atmospheric limb between tangent heights of 95 and 5 km every 65 s, during which UARS moves 4° along the orbit. Measurements occur at 1.8-s dwells between steps, and the step spacing varies from 1 km in the lower stratosphere to 5 km in the mesosphere. The FOV vertical extent at the tangent point for the 205-GHz ClO and ozone measurements is 3.5 km, the approximate vertical resolution. Horizontal resolution is ~ 400 km, set by the distance through the limb where most of the signal originates, and by orbital motion smearing during the limb scan.

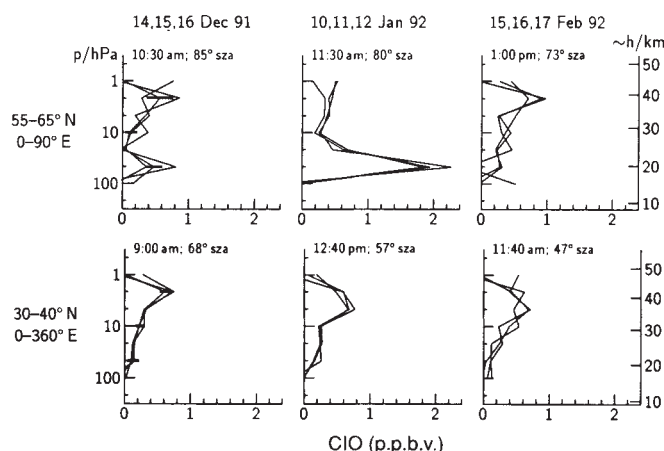


FIG. 2 CIO vertical profiles from UARS MLS. Each panel shows averages over the latitudes and longitudes indicated to the left, for each of the three days indicated at the top; ~ 12 individual profiles have been averaged for the 55–65° bins, and ~ 44 for 30–40°. Measurement local solar times and solar zenith angles (sza) for the middle day and the centre of the latitude bands are indicated. Horizontal bars in the left panels give $\pm 1\sigma$ estimated precision in the averaged profiles; absolute accuracy of $\sim 15\%$ is expected. Differences between day and night retrievals have been taken in the 30–40° averages for pressure $p \geq 10$ hPa to remove residual (~ 0.2 p.p.b.v.) systematic errors. Vertical variation is represented as piecewise-linear in logarithm of atmospheric pressure, with breakpoints at $10^{(6-n)/3}$ hPa, where $n = 0, 1, 2, \dots$, for the results reported here. Values at the breakpoints are retrieved from MLS spectral measurements taken over each complete limb scan. Approximate heights are indicated on the right.

well into darkness; less than 0.3 p.p.b.v. CIO is observed on these Q contours at ~ 195 K temperature in darkness near 0° longitude. This night-time CIO variation can be explained by a temperature dependence of ClOOC1 thermal decomposition: kinetics data³⁰ lead to night-time equilibrium CIO values of 0.7 p.p.b.v. at 210 K and 0.15 p.p.b.v. at 195 K, for 2 p.p.b.v. total chlorine in CIO and ClOOC1 at 50 hPa. Three-dimensional numerical model results for this period^{31,32} provide a more quantitative explanation of the CIO distribution seen by MLS.

Total chlorine in CIO and ClOOC1 ($Cl^* \stackrel{\text{def}}{=} \text{CIO} + 2 \times \text{ClOOC1}$), not expected to have significant diurnal variation) can be inferred³³ from CIO. We have done this for the measured CIO and coincident temperature and solar illumination (including calculated attenuation by the measured ozone, and calculated ClOOC1 thermal dissociation). CIO abundances greater than ~ 2 p.p.b.v. at 46 hPa, using nominal values³⁰ for kinetics parameters and ClOOC1 cross-sections, lead to Cl^* amounts larger than the ~ 3 p.p.b.v. inorganic chlorine in the stratosphere²¹

available for conversion to CIO. (Total atmospheric chlorine is now at 3.6 p.p.b.v., six times the natural abundance³⁴.) MLS observations in both north and south have yielded 46 hPa CIO abundances of 2–2.5 p.p.b.v., indicating virtually complete conversion of stratospheric chlorine to reactive forms, but (because of the large values of inferred Cl^*) perhaps raising questions concerning details of currently predicted CIO–ClOOC1 partitioning and ozone loss calculated from CIO. Further investigations of the large CIO values, and their implications, are in progress.

Figure 4 shows evolution of enhanced CIO on the 465 K isentropic surface during 2–13 January, the last 12 days before switch to south-viewing. The general increase seen in CIO during this period is expected because of increased ClOOC1 photolysis as the vortex edge moves to lower latitudes and the rising sun (at 60°, for example, the solar elevation angle at measurement locations increased from 2° on 2 January to 8° on 13 January, while the local time decreased from 2:30 pm to 10:30 am). The

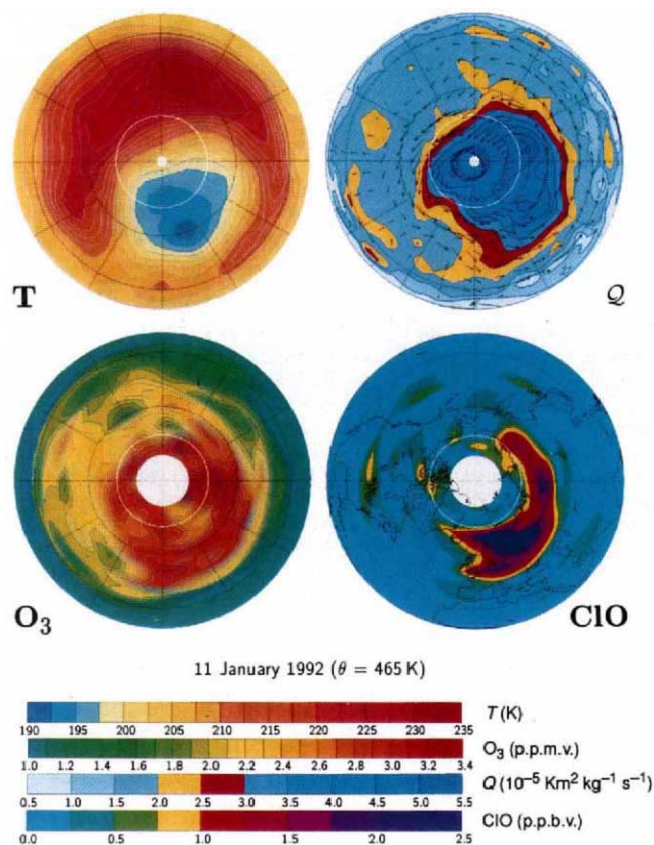


FIG. 3 The 11 January 1992 distribution of Northern Hemisphere CIO, temperature (T), ozone (O_3), and potential vorticity (Q) on the 465 K potential temperature (θ) surface. CIO and O_3 are from UARS MLS, T is from the United States National Meteorological Center (NMC), and Q is calculated from the NMC temperature and geopotential height fields. The Q map also contains calculated wind vectors whose lengths are proportional to wind speed (largest speeds are ~ 45 m s $^{-1}$, occurring near the vortex boundary). Values of Q characteristic of the vortex boundary are in red, and values of T for which type-I PSCs can form are in blue. These maps are polar orthographic projections extending outward to the Equator; thin dashed circles are at 30° and 60° latitudes, and land boundaries are included in the CIO map. Thin white circles near 70° N indicate the polar night terminator, and white areas poleward of 80° on the ozone and CIO maps indicate where no MLS measurements are obtained. Local solar times of the measurements in the CIO and ozone maps are near midday and vary only with latitude.

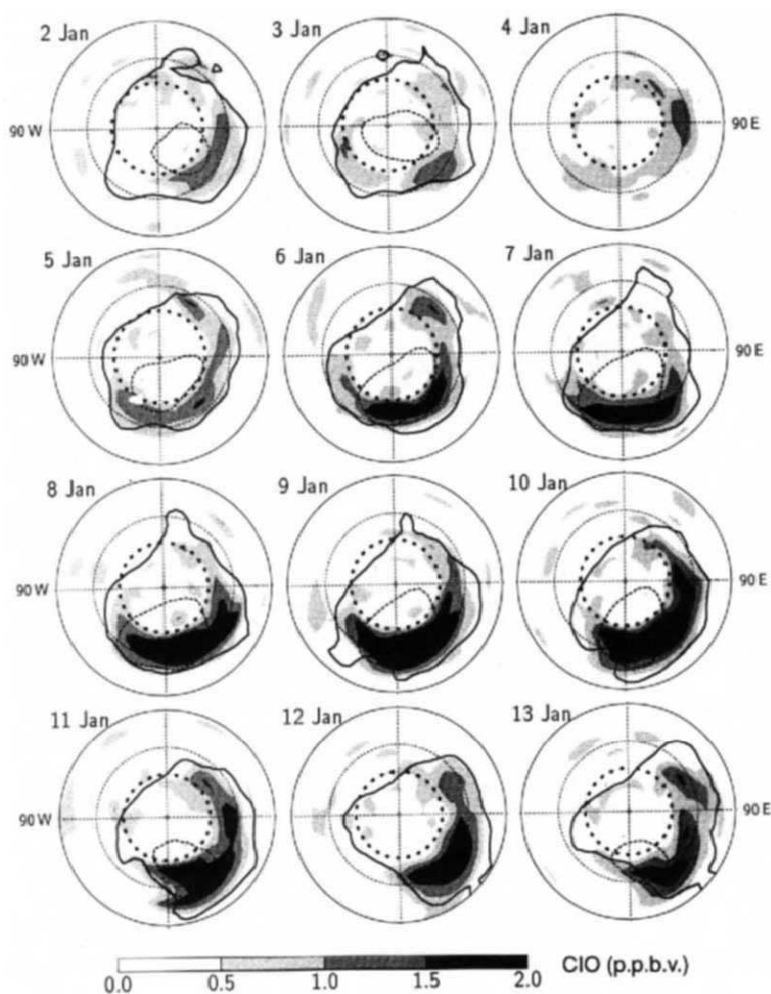


FIG. 4 Variation in the distribution of enhanced ClO during 2–13 January 1992. These maps are polar orthographic projections on the $\theta = 465$ K isentropic surface, like those in Fig. 3, but extend outward to only 40° latitude with 60° latitude indicated by the thin dashed circle. ClO abundance is indicated by the grey scale, the $2.5 \times 10^{-5} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$ potential vorticity contour by the thick solid line, the 195 K temperature contour by the thin (non-circular) dashed line, and the polar night terminator by the thick dashed circle. NMC data, missing for 4 January, showed no temperatures ≤ 195 K on 12 January at $\theta = 465$ K (but did at slightly higher altitudes, for example $\theta = 520$ K).

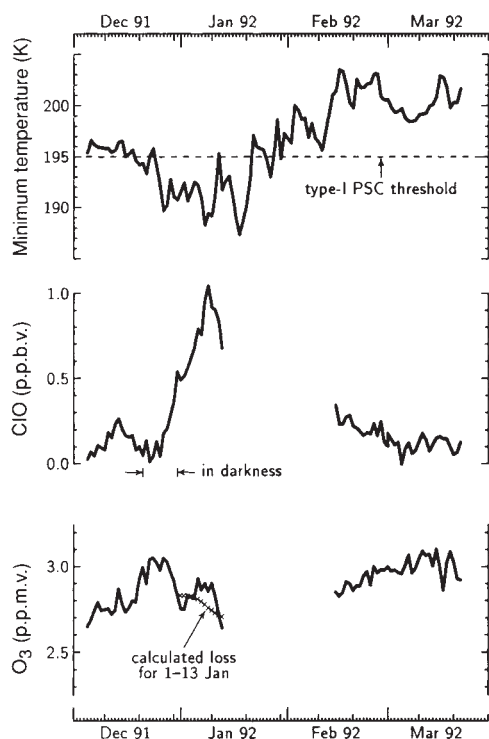


FIG. 5 Temperature, ClO and O_3 in the 1991–92 Northern Hemisphere winter vortex for a lower stratospheric layer at 465 K potential temperature. The vortex boundary is here defined as the $2.5 \times 10^{-5} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$ potential vorticity contour. Minimum temperature is given at the top, with a horizontal line indicating the approximate temperature below which nitric acid trihydrate PSCs form. Average ClO and O_3 mixing ratios, obtained by volume-integrating ClO and O_3 number densities and then dividing by the integrated air in the same volume, are given in the middle and bottom. The volume-integration includes regions of the vortex in daylight and in darkness, and the resulting ClO values are substantially lower than those for just the 'daylight' region. The bottom panel also shows expected ozone evolution with loss calculated from observed ClO and no transport into or out of the layer. Gaps in the ClO and O_3 plots between mid-January and mid-February occur when MLS was south-viewing.

pattern of enhanced ClO tends to move with Q , indicating the importance of atmospheric motions in controlling its distribution. Enhanced ClO is generally contained within $Q = 2.5 \times 10^{-5} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$, but not always (for example, 10 and 13 January, where the 195 K isotherm extended outside $Q = 2.5 \times 10^{-5} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$). Figure 4 shows that air containing large amounts of reactive chlorine can have considerable and variable meridional motion; this strongly affects ozone depletion and must be considered²⁹ for reliable calculations of ozone loss.

Ozone in the 1991–92 northern winter

The maps in Fig. 3 show correlation between ozone and Q , indicating the importance of atmospheric motions in determining the lower-stratospheric ozone distribution. It also shows that the vortex interior on 11 January contains more ozone at $\theta = 465 \text{ K}$ than the exterior, as expected from the diabatic descent³⁵ of ozone-rich air from above. The winter evolution of MLS ozone data examined on different θ -surfaces is consistent with ozone-rich mid-stratospheric air moving poleward from lower latitudes and diabatically descending in the vortex. A relative minimum in lower-stratospheric ozone developed in early January near the region of largest ClO, and on 11 January at $\theta = 465 \text{ K}$ (see Fig. 3) amounted to ~ 15 –20% relative to surrounding regions in this layer. Figure 5 shows temperature, ClO and ozone throughout the 1991–92 winter at $\theta = 465 \text{ K}$ within the Arctic vortex. The lower stratosphere was cold enough for type-I PSCs from mid-December through late January. Measured ClO and ozone are shown, integrated throughout the vortex for $Q > 2.5 \times 10^{-5} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$ and for a layer bounded by $\theta = 450 \text{ K}$ at the bottom and $\theta = 480 \text{ K}$ at the top ($\sim 2.5 \text{ km}$ vertical thickness). Vortex-integrated ClO increased greatly in early January, then decreased throughout February and March following the late-January warming above temperatures at which PSCs form. The expected climatological increase in high-latitude winter ozone is evident, but a decrease of $\sim 10\%$ occurs between late December and mid-January. Enhanced ClO was observed by MLS during late December, but these measurements occurred either in darkness or at very low solar zenith angles and an observed midday value was not obtained.

Ozone loss over time Δt due to the ClO + ClO cycle^{6,36} is given³⁷ by $\Delta O_3^c = 2k[\text{ClO}]^2[\text{M}]\Delta t \text{ molecule m}^{-3}$, where k is the rate constant³⁰ for the ClO + ClO + M reaction, M indicates a third molecule and $[X]$ is number density of species X. Using measured values of $[\text{ClO}]$, $[\text{M}]$ (obtained from temperature and pressure) and temperature (needed for k), we calculated daily fields of ΔO_3^c for 1–13 January 1992. Diurnal variation was included by inferring Cl* from ClO at each measurement point and integrating the ozone loss (expressed in terms of Cl*) over a diurnal cycle while accounting for varying ClOOCl photolysis due to changing solar zenith angle. Results from such calculations differ negligibly from those obtained by a more accurate time-dependent calculation. Estimated additional loss from the ClO + BrO cycle³⁸ was included by multiplying the calculated ΔO_3^c by 1.4, appropriate for 8 parts per 10^{12} (p.p.t.v.) BrO (previously measured³⁹ in the Arctic vortex) and 1.5 p.p.b.v. ClO (typical daytime vortex value from MLS). Localized losses up to $\sim 4\%$ per day were calculated for the 465 K layer during 1–13 January. The fields of calculated ΔO_3^c were volume-integrated in the same manner as the measured ozone and ClO, and the results are given in Fig. 5. We obtain a maximum vortex-integrated loss of $\sim 1\%$ per day for the layer, and 4% cumulative loss over the 13-day period. The measured ozone shows variations (both increases and decreases) larger than the calculated loss, illustrating the difficulty of distinguishing between chemical and dynamical effects on the observed ozone.

Comparison with the 1992 southern winter

In 1992 MLS measured ClO enhanced to >1 p.p.b.v. in the south starting at the beginning of its south-viewing period on 1 June. On this day the vortex was displaced from the pole to the

west, and ClO > 1 p.p.b.v. was observed equatorward of 60° S at 10 – 75° W , reaching 53° S at 60° W (near the Falkland Islands), and bounded on the north by $Q \approx -3.5 \times 10^{-5} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$ and on the south by the daylight edge ($\sim 65^\circ \text{ S}$) of measurements. NMC data show that the lower stratosphere had already cooled below the $\sim 195 \text{ K}$ type-I PSC threshold by 10 May, and OCIO measurements⁴⁰ at 78° S had previously indicated chlorine activation early in the 1991 southern winter. ClO increased as the winter progressed. By 14 August, when MLS started another south-viewing period, much of the lower stratosphere poleward of 60° S in daylight (equatorward of 78° S) contained ~ 2 – 2.5 p.p.b.v. ClO; the distribution at $\theta = 465 \text{ K}$ is shown on the cover. These measurements suggest that the May–August downward decadal trend¹⁰ in southern high-latitude ozone is due to chlorine but, with only one year's measurements, we cannot yet say whether 1992 might have been unusual because of effects of the Pinatubo eruption. OCIO measurements indicate increased chlorine activation by sulphate aerosol chemistry in 1992 relative to that before Pinatubo in 1991 (ref. 41). The outer boundary of enhanced ClO (at $\theta = 465 \text{ K}$) retreated poleward during late winter, to $Q \approx -5 \times 10^{-5} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$ by 20 September. Poleward retreat of the ClO boundary is under investigation and could be caused, for example, by NO_2 from the returning Sun's photolysis of HNO_3 .

There was a large loss of ozone within the Antarctic vortex in September. Figure 6 shows maps of the vertical columns of ClO and ozone for 20 September 1992, the last day of measurements before the switch to north-viewing. Corresponding maps are shown for 21 September 1991 data obtained shortly after launch. The full extent of correlation between enhanced ClO

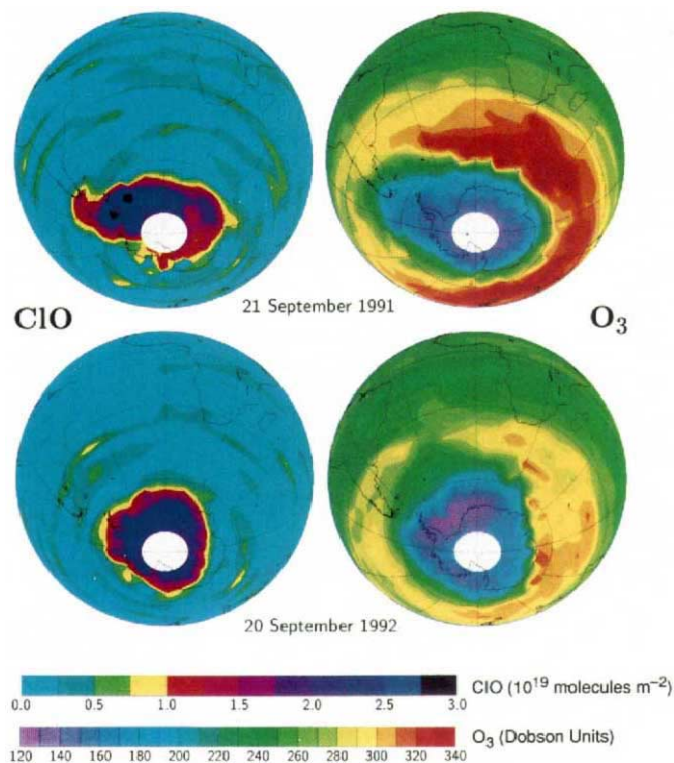


FIG. 6 ClO and O_3 in the Southern Hemisphere on 21 September 1991 and 20 September 1992. The mapped quantities are vertical columns obtained by integrating the profiles retrieved from MLS. The O_3 column is in Dobson units ($2.69 \times 10^{20} \text{ molecule m}^{-2}$) above 100 hPa. The patches of ClO outside the vortex are near the noise level. These maps were produced using linear interpolation of measurements, in contrast to those in Fig. 3, which were produced using Fourier techniques.

and depleted ozone is evident. Ozone abundances in the 'collar' around Antarctica were noticeably lower in 1992 than in 1991, and this difference is being investigated. It is informative that largest reduction in vortex ozone did not occur until September, even though greatly enhanced CIO was present from early June, with more in August than in September. This suggests a net flux of ozone into the lower stratosphere during the earlier period to counter chemical loss indicated by the enhanced CIO and, as discussed earlier for the north, our data appear consistent with diabatic descent of ozone-rich air. Continuing studies of both the horizontal and vertical evolution of the measured ozone (and other) fields should help quantify the extent to which the vortex is isolated from or interacts with the outside—crucial for implications of more widespread ozone depletion^{42,43}.

A major difference between the two hemispheres, with regard to chlorine activation and associated ozone loss, is the amount of time that the stratosphere is sufficiently cold for PSCs: this was ~5 months for the south and ~6 weeks for the north during the 1991–92 winter seasons. A related difference is minimum temperatures. At ~190 K, type-II ice PSCs can form²³ and sediment out of the stratosphere, removing HNO₃ as a source for NO₂ to quench CIO and reduce ozone destruction^{44,45} (type-I PSCs can also grow sufficiently large under suitable conditions⁴⁴ to sediment and remove HNO₃). Lower-stratospheric temperatures in the Southern Hemisphere were <190 K for ~4 months in 1992, but only occasionally so in the Northern Hemisphere 1991–92 winter. The warmer northern temperatures can be traced to stronger planetary waves resulting from greater orographic and thermal forcing by the more varied northern land distribution⁴⁶. Planetary waves can mix warm air into polar regions and induce adiabatic compression and heating, leading to diabatic cooling that results in descent across isentropic surfaces. However, planetary waves can also displace the vortex from the pole, causing more air to move into sunlight and thus increasing CIO once PSCs have formed; MLS maps have shown that this can be appreciable.

Early results from 1993 and future outlook

Minimum temperatures within the lower-stratospheric Arctic vortex in 1993 remained well below the type-I PSC threshold, and were often below the type-II threshold, until late February. This winter MLS observed ≥ 1 p.p.b.v. CIO on the $\theta = 465$ K

surface from as early as 4 December 1992 (temperatures first dropped below the type-I threshold in late November) to as late as 4 March 1993. The vortex was generally centred nearer the pole in early January 1993 than during the same period in 1992. When MLS north-viewing measurements resumed on 10 February, however, the vortex was displaced from the pole (southward to ~60° N over northeastern Canada and to ~50° N over central Russia), and ≥ 1 p.p.b.v. CIO was observed throughout most of the vortex at $\theta = 465$ K; the distribution on 14 February is shown on the cover. CIO remained high throughout February, and the vortex continued to be displaced from the pole with minimum lower-stratospheric temperatures below the type-I threshold until at least 25 February. CIO decrease occurred in March, following the late February warming, and the vortex reached as far south as 40° N over the Mediterranean. Vortex-integrated ozone in the 465 K layer decreased by 0.7% per day on average between 10 February and 1 March, whereas during the same period in 1992 (Fig. 5) it increased by 0.3% per day on average. Initial analyses of MLS data show the average ozone column (above ~200 hPa) in late winter throughout most of the extra-tropical Northern Hemisphere to be ~10% lower than in 1992, with regions that are 20% lower. The amount of ozone column decrease in northern winter observed by MLS between 1992 and 1993 is noticeably larger than the <5% winter-to-winter variations seen in the 30-year record of zonally averaged, smoothed Dobson data⁴⁷ for the north, and at the extreme of winter-to-winter variations seen in the 10-year record of zonal mean data from the Total Ozone Mapping Spectrometer (TOMS)⁴⁷. Further analyses are under way.

Industrial production of source gases that take chlorine into the stratosphere is decreasing but, even with this positive step towards protecting our ozone shield, anthropogenic chlorine in the stratosphere will continue to increase during the next decade and will be above the abundance at which the Antarctic ozone hole formed for about a century⁴⁸. Additional concerns for stratospheric ozone depletion include the detrimental effects of increasing stratospheric CH₄ (ref. 49), NO_x (ref. 50), aerosols⁵¹ and CO₂ (ref. 52). Fortunately, the technology now exists for continuous global monitoring of both the three-dimensional distribution of stratospheric ozone and processes that can deplete it. □

Received 25 January; accepted 24 March 1993.

- Molina, M. J. & Rowland, F. S. *Nature* **249**, 810–812 (1974).
- Farman, J. C., Gardiner, B. G. & Shanklin, J. D. *Nature* **315**, 207–210 (1985).
- de Zafra, R. L. *et al. Nature* **328**, 408–411 (1987).
- Anderson, J. G., Brune, W. H. & Proffitt, M. H. *J. geophys. Res.* **94**, 11,465–11,479 (1989).
- Solomon, S., Garcia, R. R., Rowland, F. S. & Wuebbles, D. J. *Nature* **321**, 755–758 (1986).
- Molina, M. J., Tso, T. L., Molina, L. T. & Wang, F. C. Y. *Science* **238**, 1253–1257 (1987).
- Solomon, S. *Nature* **347**, 347–354 (1990).
- Brune, W. H., Toohey, D. W., Anderson, J. G. & Chan, K. R. *Geophys. Res. Lett.* **17**, 505–508 (1990).
- Proffitt, M. H. *et al. Nature* **347**, 31–36 (1990).
- Stolarski, R. *et al. Science* **256**, 342–349 (1992).
- Reber, C. A. *Trans. Am. geophys. Un.* **71**, No. 51 (December 1990).
- Waters, J. W. in *Atmospheric Remote Sensing by Microwave Radiometry* (ed Janssen, M. A.) ch. 8 (Wiley, New York, 1993).
- Barath, F. T. *et al. J. geophys. Res.* (in the press).
- Jarnot, R. F. & Coffield, R. E. *The Upper Atmosphere Research Satellite (UARS) Microwave Limb Sounder (MLS) Calibration Report*, JPL Document D-9393 (Jet Propulsion Laboratory, Pasadena, California, 1991).
- Rodgers, C. D. *Rev. Geophys. Space Phys.* **14**, 609–624 (1976).
- Brasseur, G. P., Granier, C. & Walters, S. *Nature* **348**, 626–628 (1990).
- Rodriguez, J. M., Ko, M. K. W. & Sze, N. D. *Nature* **352**, 134–137 (1991).
- Brasseur, G. & Granier, C. *Science* **257**, 1239–1242 (1992).
- Tolbert, M. A., Rossi, M. J. & Golden, D. M. *Geophys. Res. Lett.* **15**, 847–850 (1988).
- King, J. C. *et al. Geophys. Res. Lett.* **18**, 2273–2276 (1991).
- Stachnik, R. A., Hardy, J. C., Tarsala, J. A., Waters, J. W. & Erickson, N. R. *Geophys. Res. Lett.* **19**, 1931–1934 (1992).
- Hanson, D. & Mauersberger, K. *Geophys. Res. Lett.* **15**, 855–858 (1988).
- McCormick, M. P. *et al. Geophys. Res. Lett.* **17**, 381–383 (1990).
- Salby, M. L. *J. Atmos. Sci.* **39**, 2577–2600 (1982).
- Lait, L. R. & Stanford, J. L. *J. Atmos. Sci.* **45**, 3784–3799 (1988).
- Clough, S. A., Grahame, N. S. & O'Neill, A. Q. *J. R. met. Soc.* **111**, 335–358 (1985).
- Randel, W. J. *J. Atmos. Sci.* **44**, 3097–3120 (1987).
- Swinbank, R. & O'Neill, A. A *Stratosphere-Troposphere Data Assimilation System*, *Climate Res. Tech. Note 35* (Hadley Centre for Climate Prediction and Research, Bracknell, 1992).
- Jones, R. L., McKenna, D. S., Poole, L. R. & Solomon, S. *J. geophys. Res.* **17**, 545–548 (1990).
- DeMore, W. B. *et al. Chemical Kinetics and Photochemical Data for Use in Stratospheric Modeling: evaluation number 10*, *Tech. Rep.* 92–20 (Jet Propulsion Laboratory, 1992).
- Kettleborough, J. *et al. Geophys. Res. Lett.* (submitted).
- Lefèvre, F. *et al. Nature* (submitted).
- Rodriguez, J. M. *et al. J. geophys. Res.* **94**, 16683–16703 (1989).
- World Meteorological Organization and United Nations Environment Program, *Scientific Assessment of Stratospheric Ozone, 1991* (WMO Geneva, 1991).
- Schoeberl, M. R., Lait, L. R., Newman, P. A. & Rosenfield, J. E. *J. geophys. Res.* **97**, 7858–7882 (1992).
- Molina, L. T. & Molina, M. J. *Phys. Chem.* **91**, 433–436 (1987).
- Anderson, J. G. *et al. J. geophys. Res.* **94**, 11480–11520 (1989).
- McElroy, M. B., Salawitch, R. J., Wofsy, S. C. & Logan, J. A. *Nature* **321**, 759–762 (1986).
- Toohey, D. W., Anderson, J. G., Brune, W. H. & Chan, K. R. *Geophys. Res. Lett.* **17**, 513–516 (1990).
- Sanders, R. W. *et al. J. geophys. Res.* (in the press).
- Solomon, S. *et al. Nature* (in the press).
- Hartmann, D. L. *et al. J. geophys. Res.* **94**, 11625–11640 (1989).
- Tuck, A. F. *J. geophys. Res.* **94**, 11687–11737 (1989).
- Toon, O. B., Turco, R. P. & Hamill, P. *Geophys. Res. Lett.* **17**, 445–448 (1990).
- Wofsy, S. C., Salawitch, R. J., Yatteau, J. H. & McElroy, M. B. *Geophys. Res. Lett.* **17**, 449–452 (1990).
- Schoeberl, M. R. & Hartmann, D. L. *Science* **251**, 46–52 (1991).
- World Meteorological Organization Rep. int. *Ozone Trends Panel 1988*, WMO Global Ozone Res. Monitoring Proj. Rep. No. 18 (WMO, Geneva, 1988).
- Benedick, R. E. *Ozone Diplomacy* (Harvard Univ. Press, Cambridge, Massachusetts, 1991).
- Blake, D. R. & Rowland, F. S. *Science* **239**, 1129–1131 (1988).
- Johnston, H. S., Kinnison, D. E. & Wuebbles, D. J. *J. geophys. Res.* **94**, 16351–16363 (1989).
- Hofmann, D. J. *Science* **248**, 996–1000 (1990).
- Austin, J., Butchart, N. & Shirre, K. P. *Nature* **360**, 221–225 (1992).

ACKNOWLEDGEMENTS. We thank many colleagues for contributing to MLS; NASA, the UARS Project and associates for UARS; R. W. Zurek's UARS team (of which G.L.M. is a member) for derived meteorological products; I. MacKenzie for diurnal integrations; W. G. Grose and team for atmospheric model data used in pre-launch simulations; A. J. Miller and A. O'Neill and their teams for meteorological data following launch; the ISAMS and CLAES UARS teams for data on polar stratospheric clouds; P. A. Newman for supplying routines that were adapted to calculate potential vorticity; D. W. Fahey, C. B. Farmer, J. R. Holton, J. A. Pyle, M. R. Schoeberl, A. F. Tuck and R. W. Zurek for comments. The work at JPL Caltech was sponsored by NASA and the work at EU by SERC.

A structural motif in the variant surface glycoproteins of *Trypanosoma brucei*

Michael L. Blum^{*†}, James A. Down^{*†}, Anne M. Gurnett[‡], Mark Carrington[§], Mervyn J. Turner[‡] & Don C. Wiley^{*||}

^{||} Howard Hughes Medical Institute, and ^{*} Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

[‡] Merck Sharpe & Dohme, Research Laboratories, Box 2000, Rahway, New Jersey 07065, USA

[§] Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK

The variable domain of the trypanosome variant surface glycoprotein (VSG) ILTat 1.24 has been shown by X-ray crystallography to resemble closely the structures of VSG MITat 1.2, despite their low sequence similarity. Specific structural features of these VSGs, including substitution of carbohydrate for an α -helix, can be found in other VSG sequences. Thus antigenic variation in trypanosomes is accomplished by sequence variation, not gross structural alteration; the extensive sequence differences among VSGs may be required for another reason, such as the avoidance of recognition by helper T cells. Additionally, VSG sequences are found to define families, within a VSG superfamily, which have evolved in the trypanosome genome.

THE African trypanosome is a parasitic protozoan that causes sleeping sickness in humans and nagana in cattle, diseases that are characterized by cyclical waves of fever. The fever cycles correlate with spikes in the population of trypanosomes in the blood¹. The trypanosome uses a large, serially expressed repertoire of antigenically distinct surface proteins to evade the immune response of its mammalian host^{2,3}. These antigens, the variant surface glycoproteins, form a coat on the surface of the parasite^{4,5} and elicit the production of VSG-specific host antibodies⁵. Some of the trypanosome population survives the host's immune response by switching to expression of one or more antigenically distinct VSGs, allowing a relapse in parasitaemia. The trypanosome genome is estimated to contain over 1,000 different VSG genes sharing little sequence homology⁶. No biochemical activity has been ascribed to these molecules, leaving the ability to fold and form a coat as the only known common requirement for their sequences.

Most VSGs consist of two domains⁷. The N-terminal 'variable' domains share low sequence homology (13–30% identity)^{8,9}, contain the antigenic epitopes accessible on the surface of the parasite¹⁰, and have recently been grouped by cysteine patterns into three classes A, B and C⁸. The C-terminal, 'conserved' domains are grouped by sequence homology into three classes, I, II and III (refs 8, 11), which appear to pair randomly with the three N-terminal classes in known VSGs. The three-dimensional structures of two class A, N-terminal domains with low sequence similarity have been determined by X-ray crystallography, namely MITat 1.2 (refs 12, 13), and ILTat 1.24 reported here¹⁴ (for VSG nomenclature, see legend to Fig. 2).

Similarities in the two three-dimensional structures of VSGs and weak amino-acid homologies, identified only by multiple sequence alignments with other class A N-terminal sequences, are here correlated to define a folding motif common to most VSGs. This analysis allows specific predictions of intra- and interchain disulphide bonds and missing secondary structure elements in other VSG sequences. It argues that trypanosomes

accomplish antigenic variation by sequence variation and not by major structural alterations. We also observe that carbohydrate can substitute for deleted polypeptide in the MITat 1.2 VSG tertiary structure. This type of substitution can be inferred in another VSG sequence, suggesting that it represents a novel structural role for oligosaccharide. The structural similarities of VSGs raise questions about the function of what appears to be a great excess of sequence diversity relative to that needed to escape antisera. They also suggest that the VSG variable domain classes (A, B, C) represent a protein superfamily developing within one genome as the result of a diversity generator, presumably evolved for antigenic variation.

Structure determination

The structure of the N-terminal domain of the ILTat 1.24 VSG was determined by X-ray crystallography. A single isomorphous heavy-atom derivative (HgI4) and its anomalous scattering provided phases to 4.2 Å resolution, which were extended to 3.8 Å resolution by solvent flattening and iterative, real-space, non-crystallographic symmetry averaging about a molecular twofold axis¹⁵. Fifty per cent of the atomic model was readily constructed, including correction of a sequence error (a 3-residue insertion). Four cycles of model building, refinement, phase combination and non-crystallographic phase averaging were required to complete the atomic model, which is currently refined at 2.7 Å resolution to an *R* factor of 20.3% with good geometry (r.m.s. $\Delta_{\text{bonds}} = 0.019$ Å, r.m.s. $\Delta_{\text{angles}} = 3.47^\circ$) (see Fig. 1 legend). An indication of the quality of the electron density maps is that three sequence errors were discovered (one insertion and two phase shifts in the nucleic acid sequence) by interpreting the electron density, and were later confirmed by resequencing⁹.

Structure of ILTat 1.24

The central elements of the ILTat 1.24 variable domain are two antiparallel α -helices (A and B in Fig. 2a) with a short turn (c) between them. From this scaffold are hung various smaller elements of secondary structure: 7 smaller α -helices (C, D, E₀, E, F, H, S) and a short, three-stranded β -sheet (near 'n' in Fig. 2). These elements are tied together by several loops (e, n, o, p, q, r, s, t, u) and one long meandering strand that covers much of the upper side (h to k) and top (k to m) of the molecule (Fig. 2, left). The N terminus is at the top of the molecule (a) and the C terminus exits at the bottom (v), presumably where it

[†] Present addresses: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA (M.L.B.); Beckton Dickinson and Company, Research Center, Box 12016, Research Triangle Park, North Carolina 27709, USA (J.A.D.).

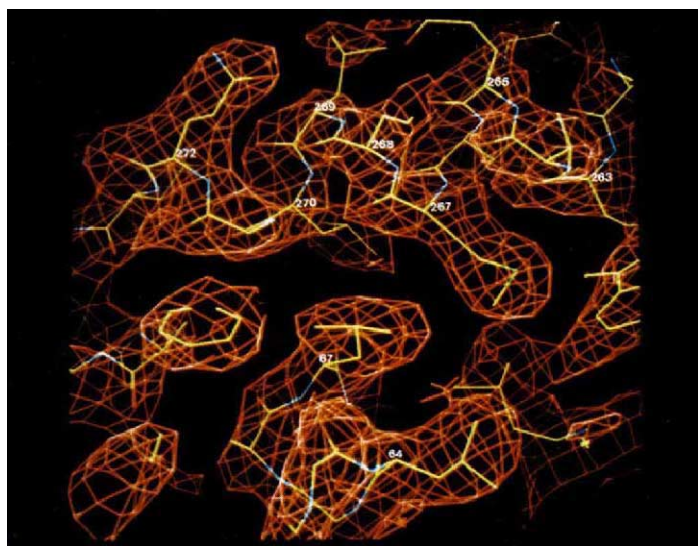
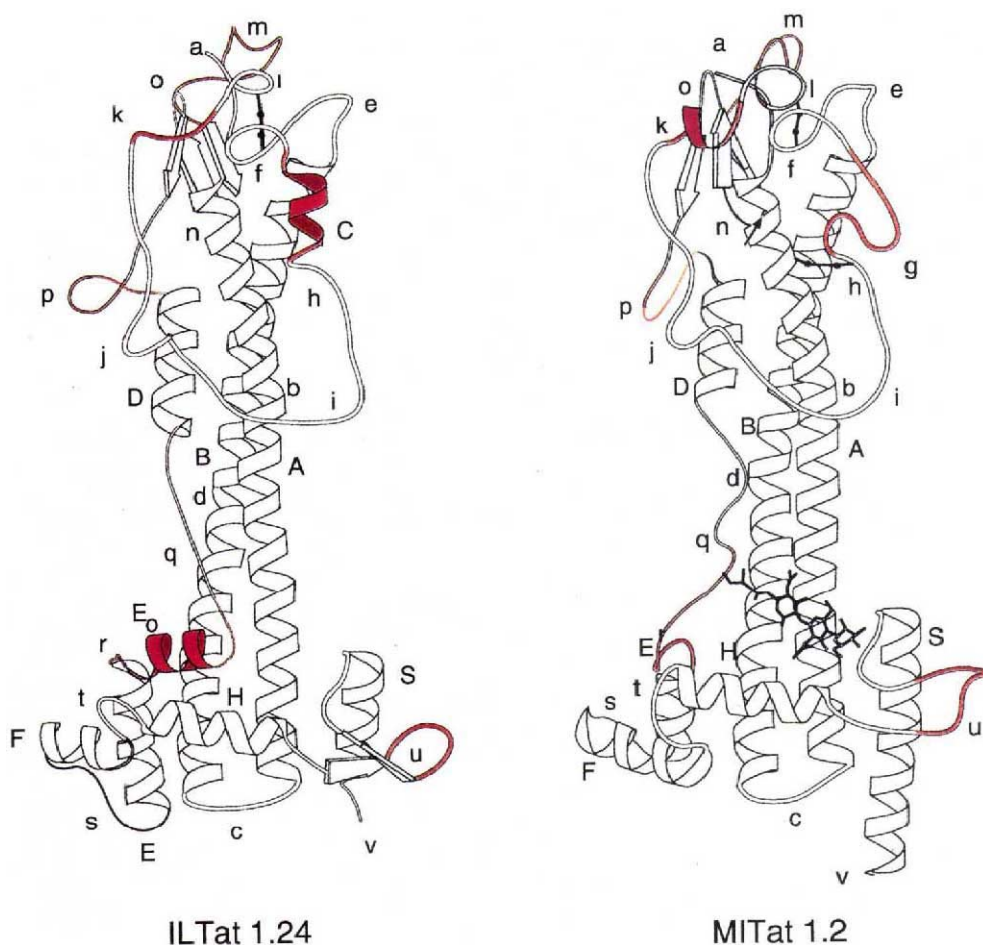


FIG. 1 Electron density map of ILTat 1.24 and the current atomic model. Electron density was calculated using current model phases and $2Fo-Fc$ terms between 10 and 2.7 Å. The view is looking up toward the E_0 helix and contacts with the B helix. Initial crystallographic phase determination and identification of the non-crystallographic symmetry (NCS) axis were carried out at 6.0 Å (ref. 47). Phases to 4.2 Å were calculated using single isomorphous replacement differences and anomalous scattering from a mercury (K_2HgI_4) derivative and improved by iterative NCS averaging and solvent flattening¹⁵. Native data to 2.7 Å were collected from seven crystals yielding 24,682 unique reflections from 107,903 independent measurements with an overall R_{sym} of 0.102. Lack of crystals prevented collection of useful high-resolution derivative data. Phases were extended to 3.8 Å by continued iteration of NCS real space averaging and phase combination. The low symmetry (twofold) prevented further extension. A partial atomic model, incorporating ~50% of the expected number of atoms, was built, using FRODO (ref. 48), into the 3.8 Å map. Subsequent refinement using CORELS and X-PLOR, phase combination with the partial model, and continued iterative application of NCS real space symmetry averaging allowed completion of an atomic model at 2.7 Å resolution. The final stages of refinement required relaxation of the NCS constraints owing to a significant bend in the molecular dimer axis. The current model consists of 5,462 atoms, including 90 H_2O molecules with mean $B=27\text{ Å}^2$, $\Delta_{bonds}=0.019\text{ Å}$ and $\Delta_{angles}=3.47^\circ$, and $R=20.3\%$ for 21,676 reflections ($F>2\sigma_F$) between 6.0 and 2.7 Å (details of the method of structure determination will be presented elsewhere). Analysis of solvent accessibility and structural congruence to MITat 1.2 presented in Fig. 4 was on the NCS constrained model¹⁴. Relaxation of this constraint does not significantly alter the discussion.

FIG. 2 Ribbon drawings^{49,50} of the two known VSG structures MITat 1.2 and ILTat 1.24. Coloured regions highlight the principal structural differences between the two molecules. Helices are identified by upper-case letters, other features are labelled with lower-case letters. a, N terminus; b, kink in A helix; c, turn between A and B helices; d, kink in B helix (MITat 1.2 only); e, loop at top of B helix; f, buried loop with disulphide bond; g, small β -sheet (MITat 1.2 only); h, loop with disulphide bond to A helix; i, large side loop; j, conserved hairpin turn; l, extended strand across top; m, dimer contact loop; n, turn one of β -sheet; o, turn two of β -sheet; p, loop at top of D helix; q, strand connecting D helix to E (MITat 1.2) or E_0 (ILTat 1.24) helix; r, E_0/E bend (ILTat 1.24) or CHO attachment site (MITat 1.2); s, E/F loop; t, F/H loop; u, long loop between H and S helices; v, C terminus. MITat 1.2, Molteno Institute *Trypanozoon* antigen type 1.2; ILTat 1.24, International Laboratory for Research on Animal Diseases, *Trypanozoon* antigen type 1.24.



enters the C-terminal domain, which was removed by an unidentified proteolytic activity before crystallization^{7,16}. ILTat 1.24 is a dimer formed by the close association of the long helices of the monomers into a 4-helix bundle, much the same as that previously determined for MITat 1.2 (ref. 13).

Structure shared with MITat 1.2

Overall, the structure of ILTat 1.24 and MITat 1.2 appear similar (Fig. 2). Using an objective measure of the similarity, which considers the spatial arrangement of corresponding residues and their orientation with respect to the preceding and succeeding residues¹⁷, 207 residues (60%) are found to form the same structure. This shared core of structure, coloured in Fig. 3, is comprised of the two long helices coloured blue, five smaller helices that decorate the central helices coloured red, parts of three β -strands and a buried turn between two strands coloured orange, and five short segments of loop coloured yellow. The 207 residues of common structure share only 16% sequence identity (Fig. 4) and have an r.m.s. deviation in $C\alpha$ positions of 1.8 Å, similar to that found in other families of protein structure¹⁸. When aligned by sequence alone, the entire N-terminal domain sequences of MITat 1.2 and ILTat 1.24 show a higher (20%) sequence identity, but this increased similarity is an artefact of sequence alignment unconstrained by knowledge of tertiary structure. The remaining 40% of the structure of the two VSGs is not conserved in detail, but within this portion there are only six segments of significantly different structure (coloured red in Fig. 2). Two α -helices in ILTat 1.24 (C and E₀) are strands in MITat 1.2, one α -helix in MITat 1.2(k) is a strand in ILTat 1.24. Three loops (m, p, u) also have different conformations.

One region of different structure is particularly interesting because it includes the single N-linked glycosylation site in the MITat 1.2 variable domain sequence. The first three monosaccharide residues at this site were sufficiently well ordered in the MITat 1.2 electron density map that they could be modelled¹⁴. The E₀ helix in ILTat 1.24 is found to occupy about the same volume in space as that occupied by carbohydrate in MITat 1.2 (Fig. 2). Amino-acid positions in the conserved core that contact monosaccharide residues in MITat 1.2 contact amino-acid residues of the E₀ helix in ILTat 1.24 (not shown). Thus an oligosaccharide in MITat 1.2 appears to substitute for a helix in ILTat 1.24, each burying residues of the conserved core.

Domain class evident in complete sequence

Although sequence homologies have been noted in the first 30 amino-acid residues of VSGs¹⁹, analysis of complete sequences has generally not identified a basis for a common VSG structure²⁰⁻²². A recent attempt at aligning all of the known complete VSG sequences²² used pairwise sequence alignments, a method known to be inadequate on sequences of such low homology²³. In the absence of added structural information, such alignments are no more likely to be correct than many other possibilities²³. (The alignments in question²² even failed to match conserved cysteine positions.) The following analysis corrects problems with earlier alignment attempts in light of the realization that differences among all VSGs are not uniformly distributed and that VSG sequences fall into identifiable families. The known three-dimensional structures of two VSGs provide powerful clues for verification of the new alignments and provide a context in which to evaluate the role of observed sequence homologies. Recent identification of variable domain classes based on cysteine positions⁸ suggested that sequence analysis might benefit from prior segregation based on this classification. To test this, twenty-three VSG variable domain sequences were aligned, pairwise, using a standard algorithm²⁴ and a protein similarity scoring matrix²⁵. When the sequences are divided into classes (A, B, C), they are found to share significant homology among members of a class, and to show no statistically significant homology between classes. The range of individual scores was

low and wide (from -1.6 to 26.7 σ , mean of 2.8), but a sequence could be classified by score alone when its mean score was calculated for all sequences of each class (Table 1). This indicates that classes are evident in the entire sequence, not only in the cysteine positions. The sequences, segregated by class and then aligned within each class, further indicate that a majority of secondary and tertiary structure must be conserved within classes.

Structural motif in class A sequences

Segregation of the VSG sequences by class greatly facilitated alignment (Fig. 4) and further suggested that features of the aligned set of sequences might be correlated with structural features found to be conserved between MITat 1.2 and ILTat 1.24. In an alignment of ten class A, VSG variable domain sequences, including MITat 1.2 and ILTat 1.24 (Fig. 4), 81 positions were found to exhibit conservation of the character of the amino-acid residue across eight of the ten sequences (blue

TABLE 1 Mean alignment scores for 21 VSG N-terminal sequences when aligned pairwise with all other VSG N-terminal sequences

Variant	All As	All Bs	All Cs
MVAT 5	5.7	0.8	1.8
AnTat 1.10	5.6	0.9	2.3
ILTat 1.3	7.0	1.0	0.2
MVAT 4	4.4	0.7	1.2
MITat 1.1	8.1	1.0	2.1
MITat 1.4	6.7	0.7	2.6
MITat 1.5	4.1	1.1	0.4
MITat 1.6	7.0	1.1	1.1
ILTat 1.22	1.7	0.0	3.2
ILTat 1.24	7.1	1.0	1.8
MITat 1.2	8.5	1.1	3.2
All class A	6.0	0.9	1.8
ILTat 1.21	0.4	7.1	0.6
ILTat 1.23	1.1	4.4	1.5
ILTat 1.25	1.8	5.3	0.8
WRATat A	1.8	7.3	0.2
WRATat B	0.5	6.5	0.5
YNat 1.1	0.3	4.5	0.4
YNat 1.3	0.0	5.6	-1.1
All class B		5.8	0.4
BoTat 20	2.4	-0.4	10.1
ILTat 1.1	1.8	0.3	11.5
MVAT 7	1.3	1.3	12.4
All class C			11.4

The program ALIGN^{24,40} and a similarity scoring matrix²⁵ were used. The true sequence length of the N-terminal domain was approximated by truncating the sequence one residue short of the first cysteine residue recognized as belonging to the C-terminal homology domain. Scores are expressed as number of standard deviations above the mean score for 50 alignments of randomized sequences. The sequences represent all the published sequences (as of November 1991), excluding those that are very closely related (for example, AnTat 1.1, which is 70% homologous to AnTat 1.10). Each VSG sequence, when aligned pairwise with all other sequences, scores best against other sequences of its class. ILTat 1.22 is an exception, scoring better against the class C sequences than against the other As. The cysteine positions in ILTat 1.22 do not appear to be related to the class C cysteine positions; it is possible that it is either a hybrid or a member of new class. The intraclass scores are significantly greater (by 4 to 12 σ) than random. The interclass scores are only barely (1 to 2 σ) significant. The mean score was 3.0 for all 210 possible pairwise alignments. The sequences were classified as A, B or C based on the pattern of cysteines⁸. Class A: MVAT 5 (ref. 41), AnTat 1.10 (ref. 42), ILTat 1.3 (ref. 9), MVAT 4 (ref. 43), MITat 1.1 (ref. 8), MITat 1.4 (ref. 44), MITat 1.5, MITat 1.6, ILTat 1.22, ILTat 1.24, and MITat 1.2 (ref. 8); class B: ILTat 1.21, ILTat 1.23 and ILTat 1.25 (ref. 8), WRATat A, and WRATat B (ref. 41), YNat 1.1 and YNat 1.3 (ref. 21); class C: BoTat 20 (ref. 45), ILTat 1.1 (ref. 46), MVAT 7 (ref. 43).

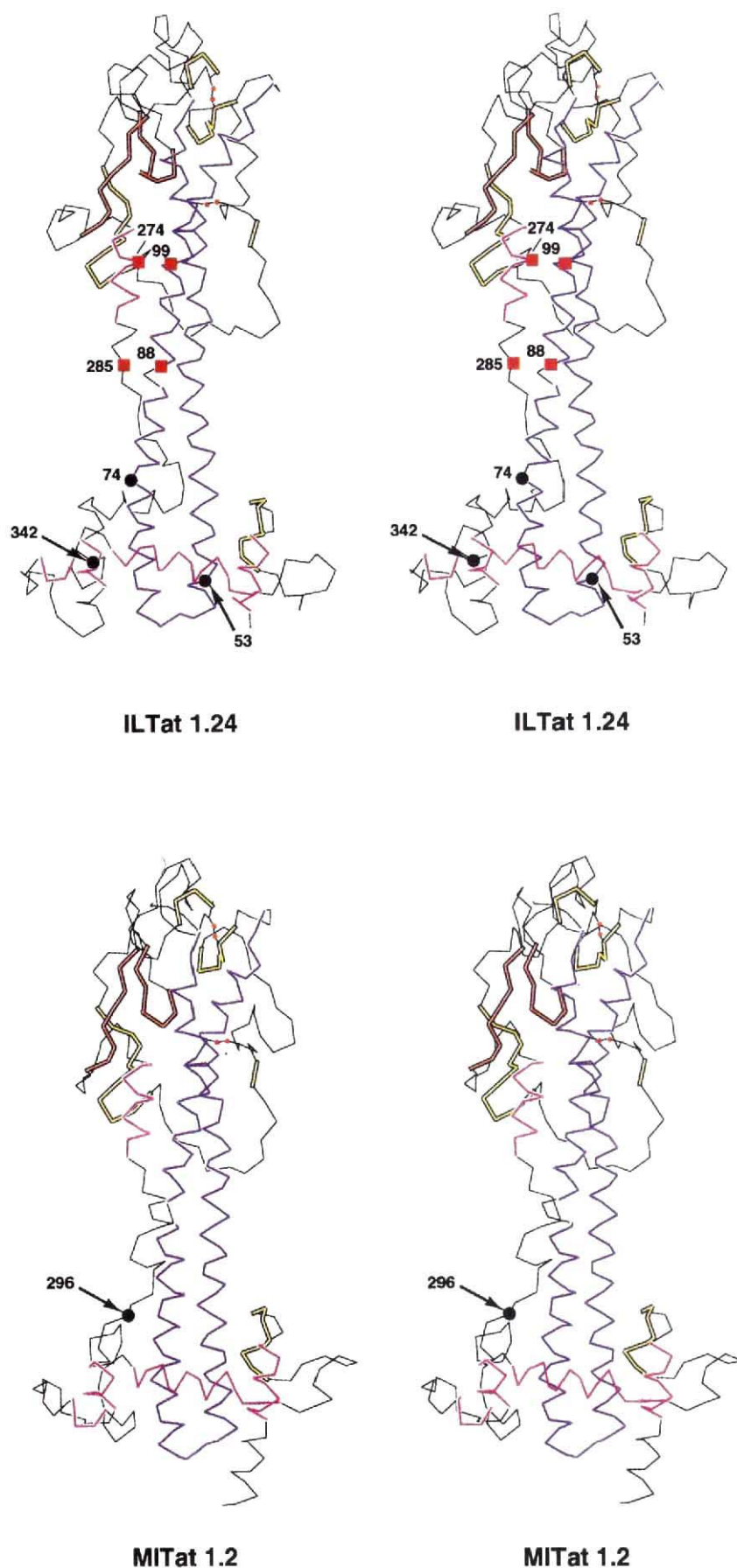


FIG. 3 Stereo drawings of trace through C α positions of the two known VSG structures. The conserved core structure (see text) is coloured to correspond to sequence alignment in Fig. 4. Dark blue, central helical core; yellow, various conserved loops; orange, β -sheet region; pink, smaller conserved helices. Top, ILTat 1.24. Black dots show where glycosylation sites in MITat 1.5 map to the ILTat 1.24 structure based on sequence alignment (Fig. 4). These sites lie such that carbohydrate at 53 and 345 and at the dimer-related 74 would be positioned where they might substitute for parts of the H or S helix (see text). Bottom, MITat 1.2. The observed glycosylation site at 296 in MITat 1.2 is marked. Note that sequence numbers refer to the consensus sequence of Fig. 4 and not to the individual molecular sequences. Red squares indicate the positions of cysteines in the sequences of MVAT 4 (99 and 274) and MITat 1.6 (88 and 285) based on the sequence alignment. Each of these pairs is predicted, by their proximity in the folded structure, to form a disulphide bond.

columns in Fig. 4). Three-quarters of these positions are buried (inaccessible to solvent) in both of the VSG structures (unpublished data), which suggests that their character has been conserved to maintain intramolecular contacts that stabilize the VSG fold. Most significantly, the positions of the eighty-one conserved amino acids cluster in regions (black blocks, Fig. 4) where MITat 1.2 and ILTat 1.24 share common structure (coloured rows in Fig. 4) and display patterns indicative of the observed secondary and tertiary structure.

Specific features of the shared core structure are evident in

the pattern of conservation in the aligned, class A sequences. Quasi-periodic repeats²⁶ suggest the presence of the A, B and D helices (dark blue rows in Fig. 4) and conserved residue character is present in all the other helices (violet blocks in Fig. 4), except F, which has only 1.5 α -helical turns conserved between ILTat 1.24 and MITat 1.2. Conserved residue character is also present in the three-stranded β -sheet (orange blocks, Fig. 4), and four of the five conserved loop segments (yellow blocks, Fig. 4). Regions of conserved glycine or proline (green columns, Fig. 4) define several turns; appearing between the two long

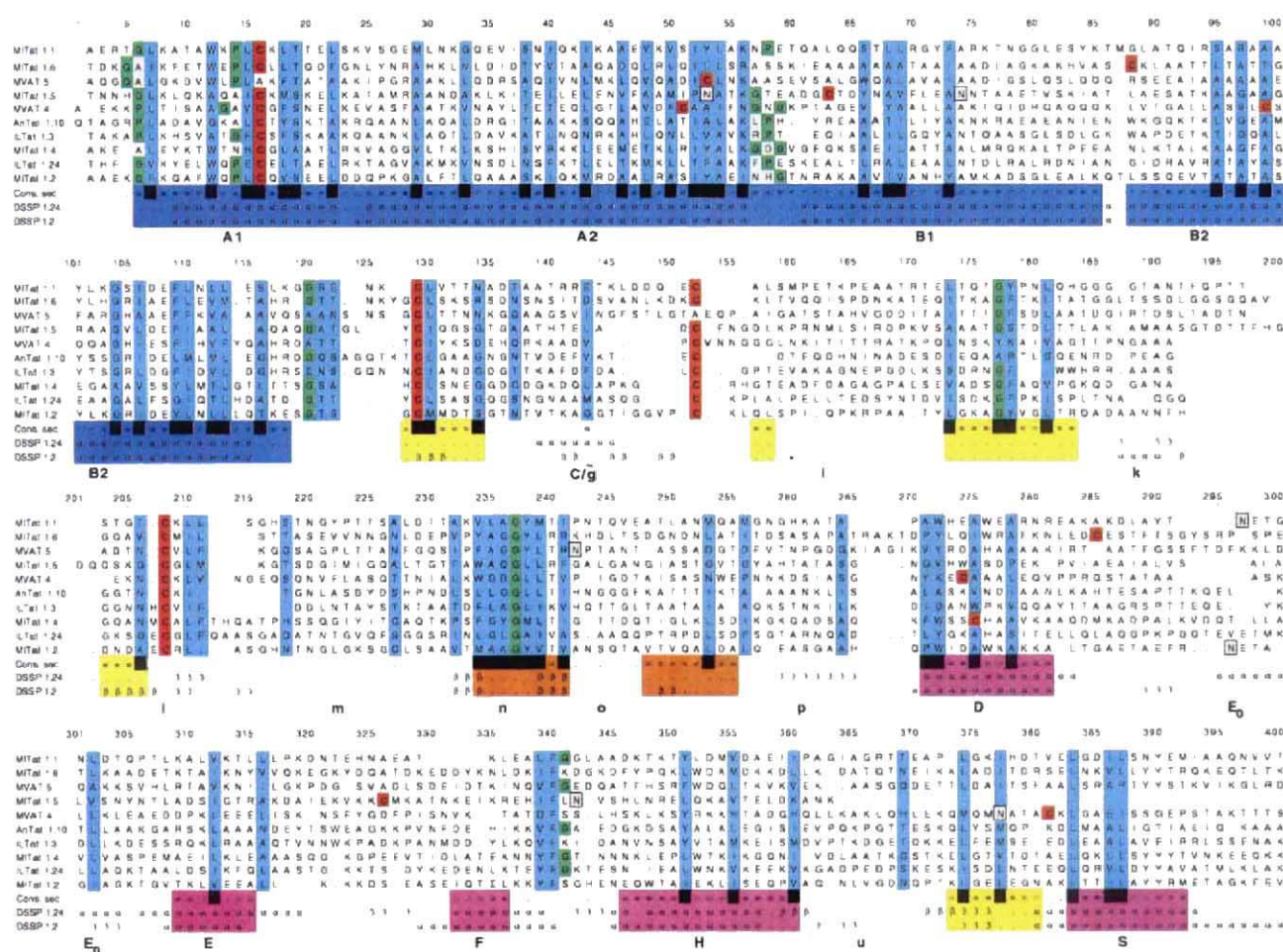


FIG. 4 VSG sequences alignments. The aligned sequences of 10 class A VSG variable domains showing the relationship of conserved structural features of ILTat 1.24 and MITat 1.2 to identifiable homologies in the aligned sequence set. Alignment was accomplished in two steps. First, all 10 of the sequences were aligned with an n -dimensional alignment computer program (TULLA⁵¹) using the protein similarity scoring matrix²⁵. Then, the sequences were adjusted manually (as recommended by S. Subbiah, personal communication) using a sequence editor (LINEUP⁵²) in order to align recognizable features, such as the region between 232 and 241, that the program did not successfully align across all sequences. The failure of the program to correctly align these features was largely due to large deletions in nearby regions in one or more sequences. Extensive manual adjustment was also necessary to reconcile the alignment with the known three-dimensional structural alignment between ILTat 1.24 and MITat 1.2. The alignment of these two sequences, presented here, is based on the known three-dimensional structures. Adjustment of the alignment using knowledge of the three-dimensional structures²³ was crucial to recognizing structural features which are indeed present in the properly aligned sequences, but are hidden by the low sequence homology. The use of structural information identified the conserved β -sheet region (232–241), and the hydrophobic conservation in helix H, that have been missed in earlier alignment attempts²². Three lines below the sequences indicate structural properties: Cons. Sec., an 'equals' sign is shown at each position where ILTat 1.24 and

MITat 1.2 have similar secondary structure as determined by OVLAP¹⁷. DSSP 1.24 and DSSP 1.2, secondary structure designation as determined by the program DSSP⁵³ for ILTat 1.24 and MITat 1.2 respectively where α indicates α -helix, β indicates β -sheet and 3 indicates 3/10-helix. Horizontal coloured blocks identify regions of conserved structure between ILTat 1.24 and MITat 1.2 and correspond to the colour coding of Fig. 2. Sequence alignment is highlighted as follows: light blue, homologous residue in 8 of the 10 sequences; red, cysteines; green, glycine or proline in positions defining conserved turns; boxed residues are potential N -linked glycosylation sites. Black rectangles indicate positions where class A sequence homology falls in a region of observed structure conservation. Bold faced lettering below the alignment indicates structural features as labeled in Fig. 2. Homology is defined liberally as hydrophobic, Y=W=F=M=L=I=V=A=G; polar, N=Q=T=S=G; positive charged, K=R; and negative charged, E=D. The figure is truncated at position 400 (out of 425) due to space considerations. An alignment of class B sequences, where no X-ray structure is yet known, is available from the authors. It suggests that class B VSGs contain two disulphides homologous to those conserved in class A sequences. The predicted disulphides are, for example, ILTat 1.25 cysteine 177 to 240 and 16 to 198 or 200. Because of the low sequence homology, the alignments of any one sequence in both class A and more so in class B lists may still contain errors in the details, but such errors are not expected to change the overall picture that emerges from them collectively.

helices (57–59), at the end of helix B (120), at a reverse turn in segment j (177), at the turn between strand 1 and 2 of the three-stranded β -sheet (237), and between helix F and H (341). A conserved glycine at position 104 (Fig. 4) defines a position in the VSG dimer interface where two B2 helices are able to make their closest approach¹³.

The four cysteine residues present in both MITat 1.2 and ILTat 1.24 (positions 16 and 152, and positions 129 and 209 in Fig. 4), are the defining characteristic of class A sequences⁸, and form two disulphide bonds which are known to be conserved in the two structures (Figs 2 and 3) and in MITat 1.4 (ref. 27). The MVAT 5 sequence lacks both members of the 16,152 disulphide pair (Fig. 4), but fits the remainder of the conserved core pattern (blue columns in Fig. 4), arguing that the disulphide bonds are not absolutely required to maintain the folding motif. The bonding of additional cysteines present in some of the class A sequences can be inferred when the sequences are folded into the class A motif, and provide strong evidence that the alignment is valid. Both MVAT 4 and MITat 1.6 have pairs of cysteines (99, 274 and 88, 285 respectively) which are predicted by the alignment to occur in ideal juxtaposition to form disulphide bonds (Fig. 3). Antat 1.1 and 1.10 both have a single, odd cysteine that maps to the end of the S helix (position 410; not shown in Fig. 4) where a disulphide could form between monomers in the dimer and both molecules have been observed to occur as disulphide-linked dimers²⁸. MITat 1.4 also has an odd cysteine but at 275, a location mapping to helix D, that could not possibly form a symmetric, inter-monomeric disulphide bond. That cysteine has been observed to be a free sulphhydryl²⁷.

Carbohydrate can be inferred to substitute for a region of polypeptide in the MITat 1.5 class A variable domain. Six of the seven potential N-linked oligosaccharide attachment sites in the sequences in Fig. 4 appear from the alignment with the known structures to be on the surface of the VSG. MITat 1.5 contains three glycosylation sites (53, 74, and 342) but one site (53) would be buried by contacts with the H helix (Fig. 3, top). In the aligned sequences (Fig. 4), the H and S helices are characterized by periodic homologous residues (blue columns) coinciding with buried positions in both MITat 1.2 and ILTat 1.24, a characteristic of amphipathic helices with a hydrophobic face to the interior. The A class VSGs all share these features, except MITat 1.5, which has a shorter sequence apparently lacking all, or most, of the H and S α -helices. Residues 52 and 69 on the A and B α -helices are conserved hydrophobic positions buried by the H and S α -helices. The cluster of three oligosaccharide sites in MITat 1.5 (Fig. 3, top, black dots) facing toward the volume occupied by the H and S helices in the VSG dimer suggests that carbohydrate may bury these conserved core residues in the MITat 1.5 structure, effectively substituting for polypeptide just as carbohydrate in MITat 1.2 appears to replace the E₀ α -helix in ILTat 1.24 (Fig. 2).

Other VSG classes

The patterns of conservation of cysteines and of sequence homology, particularly of hydrophobic residues, in other VSG sequence classes can be used to extend structural arguments to those classes. The variable domain class B currently has a sufficient number of known sequences that patterns of conserved homology can be identified. All known VSG sequences in class A, B and C (except MVAT 5) have a cysteine near position 13–16. This in itself makes some conservation of tertiary structure likely but no evidence is yet available to indicate the oxidation state or possible disulphide partner of this cysteine in any but the class A proteins. Comparison of a set of seven class B sequences (data not shown, but available from authors) reveals that these sequences have a pattern of conserved residue type in the region surrounding the first cysteine that is similar to the pattern observed in class A (Fig. 4). The class B sequences also have in common with class A sequences, two long regions

of reduced Pro and Gly content, with hydrophobic heptad repeats, separated by a four-position region of high Pro and Gly content. This indicates that the class B structures are based on two long antiparallel α -helices connected by a turn as observed in the class As, although the length of the class B α -helices and the exact form of the turn may vary. The class B sequence alignment also shows several highly conserved hydrophobic heptads near the C-terminal end of the variable domain (not shown), indicating 3 or 4 medium length helices as in the class A structure, and a few conserved glycine positions indicating conserved turns. As yet, we have not succeeded in matching these features to specific features in the class A structures, although the alignment does suggest a probable disulphide bonding pattern resembling the class A pattern (Fig. 4 legend). There are not yet enough sequences available to analyse the class C VSGs.

Discussion

Our structural studies of ILTat 1.24 and MITat 1.2 VSGs demonstrate that the African trypanosomal antigens accomplish antigenic variation through variation of sequence and limited conformational modification and not by gross alteration of structure. The two structures are essentially superimposable over 60% of their residues, although their amino-acid sequences share only 16% identities in those regions. Using an *n*-dimensional alignment algorithm supplemented with manual adjustments and knowledge of the structural overlap of ILTat 1.24 and MITat 1.2, ten class-A VSG sequences were aligned, making it possible to identify a set of positions where amino-acid character (but not identity) is conserved. Most of these positions are buried in the known structures and therefore contribute to the stability of the VSG fold. The set of these conserved positions overlaps almost all the shared features of secondary and tertiary structure in the two known structures. This argues that all of the class A sequences have elements of this common three-dimensional structure, presumably with some differences like those seen between MITat 1.2 and ILTat 1.24 (Fig. 2). From the sequence alignment, specific details of the structure of unknown VSGs have been inferred, for example the positions of intra- and interchain disulphide bonds, the location of oligosaccharide sites and, in some cases, the absence of elements of structure. A novel role for carbohydrate may have been discovered: an oligosaccharide is observed in MITat 1.2 to substitute for the deletion of an α -helix in ILTat 1.24 by burying the same core residues that the polypeptide had buried; a similar substitution has been inferred from the sequence of MITat 1.5, in which a cluster of three oligosaccharides appear to be positioned to replace part of a surface structural element consisting of two helices and a loop between them that seems to be deleted in that sequence.

The diversity present in VSG sequences appears to be much greater than is required to alter antibody-binding sites on VSGs in order to escape neutralization. In the case of viruses like influenza and rhinovirus, antigenic variation sufficient to escape neutralization from antibodies raised in previous infections requires only about a dozen point mutations on the surface of exposed proteins^{29,30}. Point mutations in VSGs also alter antibody binding to exposed epitopes³¹. Yet essentially the entire VSG sequence, comprising most of the surface and most of the interior of the N-terminal domain of the protein is changed between trypanosome antigenic variants. Why does the trypanosome VSG exhibits such extreme sequence variation? Two answers seem plausible. First, the trypanosomes' diversity generator^{32–34} may simply be so powerful that it has generated more diversity than necessary to escape neutralization. Excess sequence diversity, beyond that needed for antigenic variation, may simply be accommodated to the limit that will still fold into a VSG structure and form a cellular coat. A second possibility is that the trypanosome requires the VSG sequence diversity to escape recognition both by antisera and by helper T cells.

Helper T cells recognize short peptides processed from antigens and presented on host major histocompatibility complex (MHC) molecules. To escape helper T-cell recognition might require variation throughout the VSG antigen's sequence in order to alter all potential peptides. Effective T-cell help during a viral infection seems to be short-lived, peaking in mice within 3 to 9 days of initial infection³⁵. Re-infection within 4 or 8 days with a serologically distinct viral strain, which nevertheless shares helper epitopes, is rapidly neutralized, but re-infection after ≥ 15 days is not. Re-infection by new virus strains of previously infected individuals often occurs only after weeks, months or years, so that antigenically distinct viral strains would need only to be serologically distinct and not variable at the level of T-helper epitopes (HIV-1 infection may be an exception³⁶). The timing of a secondary trypanosome infection by an antigenic variant in trypanosomiasis, however, occurs within several days of the first infection¹. Trypanosomes may require variation in the VSG helper epitopes, achieved by changes in residues over the entire VSG sequence in order to avoid the T-cell help available within the effective lifetime of 4–8 days. T-cell help plays a part in increased antibody response to VSGs³⁷, but the nature and importance of the helper epitopes during successive waves of parasitaemia has not been established.

Structural arguments suggests that the three classes of N-terminal VSG sequences, A, B and C, comprising the thousand VSG genes in a trypanosome, may represent a developing protein superfamily. Only one N-terminal proximal cysteine is common to all sequences, but where a number of sequences can be aligned (class B), periodic conservation of mostly hydrophobic residues

divided by a glycine-rich turn region argues for the presence of the two long, antiparallel α -helices (A and B) found at the core of the class A structures. Determination of the disulphide bond pattern or the three-dimensional structure of class B and C VSGs may be required to further define the relationship (a prediction is given in Fig. 4 legend). But if, as seems likely, all VSGs have recognizably similar structures, the disparate VSG classes, like the antibodies, CD4, CD8, MHC antigens, T-cell receptors and ICAMs of the immunoglobulin superfamily³⁸, may represent a protein superfamily whose membership is more difficult to diagnose from sequence alone. Evolution of a protein superfamily evidently requires the duplication of genes and subsequent (segregated) divergence, but in trypanosomes even the duplicated copies of genes appear to be linked through gene conversion events³⁹. Class A, B and C genes may be able to mix together in diversity-generating recombinations (or gene conversions), but A, B and C classes have sufficiently diverged that proteins resulting from mixing classes would either not fold stably or not form VSG surface coats. Alternatively, the groups of A, B and C genes may have become segregated by some genetic mechanism so that the diversity-generating mechanism does not mix them. In either case, if sequenced, the thousand VSG genes may offer the opportunity in one genome for analysing the evolution of superfamily-level diversity in a protein fold.

Whatever the cause, the extreme sequence variation of VSGs observable within the trypanosome genome may provide important clues about what sequence constraints are imposed by a particular protein fold. The ability to align the VSGs should greatly facilitate this study of VSG structure. □

Received 9 March; accepted 16 March 1993.

- Ross, R., & Thomson, D. *Proc. R. Soc. Lond.* **B82**, 411–415 (1910).
- Donelson, J. E. *Contrib. Microbiol. Immun.* **8**, 138–175 (1987).
- Cross, G. A. M. *A. Rev. Immun.* **8**, 83–110 (1990).
- Vickerman, K. *J. Cell Sci.* **5**, 163–193 (1969).
- Cross, G. A. M. *Parasitology* **71**, 393–417 (1975).
- Van der Ploeg, L. H. T. *et al. Nucleic Acids Res.* **10**, 5905–5923 (1982).
- Johnson, J. G. & Cross, G. A. M. *Biochem. J.* **178**, 689–697 (1979).
- Carrington, M. *et al. J. molec. Biol.* **221**, 823–835 (1991).
- Rice-Ficht, A. C., Chen, K. K. & Donelson, J. E. *Nature* **294**, 53–57 (1981).
- Miller, E. N., Allan, L. M. & Turner, M. J. *Molec. biochem. Parasit.* **13**, 309–322 (1984).
- Cross, G. A. M. *J. cell. Biochem.* **24**, 79–90 (1984).
- Freyman, D. M. thesis, Harvard Univ. (1988).
- Freyman, D., *et al. J. molec. Biol.* **216**, 141–160 (1990).
- Blum, M. L. thesis, Harvard Univ. (1990).
- Bricogne, G. *Acta crystallogr.* **A32**, 832–847 (1976).
- Metcalfe, P., Down, J. A., Turner, M. J. & Wiley, D. C. *J. biol. Chem.* **263**, 17030–17033 (1988).
- Rossmann, M. G. & Argos, P. *J. molec. Biol.* **105**, 75–95 (1976).
- Chothia, C. & Lesk, A. M. *EMBO J.* **5**, 823–826 (1986).
- Olafson, R. W. *et al. Molec. biochem. Parasit.* **12**, 287–298 (1984).
- Lalor, T. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 998–1002 (1984).
- Strickler, J. E. *et al. Biochemistry* **26**, 796–805 (1987).
- Reinitz, D. M., Aizenstein, B. D. & Mansfield, J. M. *Molec. biochem. Parasit.* **51**, 119–132 (1992).
- Lesk, A. M., Levitt, M. & Chothia, C. *Prot. Engng* **1**, 77–78 (1986).
- Needleman, S. B. & Wunsch, C. D. *J. molec. Biol.* **48**, 443–453 (1970).
- McLachlan, A. D. *J. molec. Biol.* **61**, 409–424 (1971).
- Cohen, C. *et al. Nature* **311**, 169–171 (1984).
- Allen, G. & Gurnett, L. P. *Biochem. J.* **209**, 481–487 (1983).
- Hublart, M., Mendonça-Previato, L., Boutignon, F., Huet-Duvillier, G. & Degand, P. *Comp. Biochem. Physiol.* **92B**, 705–710 (1989).
- Wiley, D. C. & Skehel, J. J. *A. Rev. Biochem.* **56**, 365–394 (1987).
- Rossmann, M. G. *et al. Nature* **317**, 145–153 (1985).
- Baltz, T. *et al. EMBO J.* **10**, 1653–1659 (1991).
- Borst, P. & Greaves, D. R. *Science* **235**, 658–667 (1987).
- Frasch, A. C. C., Borst, P. & Van den Burg, J. *Gene* **17**, 197–211 (1982).
- Guyaux, M., Cornelissen, A. W. C. A., Pays, E., Steinert, M. & Borst, P. *EMBO J.* **4**, 995–998 (1985).
- Roost, H.-P., Charan, S. & Zirkelagel, R. M. *Eur. J. Immun.* **20**, 2547–2554 (1990).
- Phillips, E. P. *et al. Nature* **354**, 453–459 (1991).
- Reinitz, D. M. & Mansfield, J. M. *Infect. Immun.* **58**, 2337–2342 (1990).
- Lesk, A. M. & Chothia, C. *J. molec. Biol.* **160**, 325–342 (1982).
- Donelson, J. E. & Rice-Ficht, A. C. *Microbiol. Rev.* **49**, 107–125 (1985).
- Orcutt, B. C., Dayhoff, M. O., George, D. G. & Barker, W. C. *ALIGN* (Protein Identification Resource, National Biomedical Research Foundation, Georgetown University Medical Center, Washington DC, 1986).
- Reddy, L. V., Hall, T. & Donelson, J. E. *Biochem. biophys. Res. Commun.* **169**, 730–736 (1990).
- Pays, E. *et al. Cell* **34**, 371–381 (1983).
- Lenardo, M. J., Rice-Ficht, A. C., Kelly, G., Esser, K. M. & Donelson, J. E. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6642–6646 (1984).
- Boothroyd, J. C., Paynter, C. A., Coleman, S. L. & Cross, G. A. M. *J. molec. Biol.* **157**, 547–556 (1982).
- Thon, G., Baltz, T. & Eisen, H. *Genes Dev.* **3**, 1247–1254 (1989).
- Rice-Ficht, A. C., Chen, K. K. & Donelson, J. E. *Nature* **298**, 676–679 (1982).
- Metcalfe, P., Blum, M., Freymann, D., Turner, M. & Wiley, D. C. *Nature* **325**, 84–86 (1987).
- Jones, T. A. *Meth. Enzym.* **115**, 157–171 (1985).
- Richardson, J. S. *Adv. Prot. Chem.* **34**, 167–339 (1981).
- Priestle, J. P. *J. appl. Crystallogr.* **21**, 572–576 (1988).
- Subbiah, S. & Harrison, S. C. *J. molec. Biol.* **209**, 539–548 (1989).
- Devereux, J., Haeblerli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).
- Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).

ACKNOWLEDGEMENTS. We acknowledge the earlier contributions of P. Metcalfe and D. Freymann. This work was supported by the NIH (D.C.W.). J.A.D. was supported by a Fellowship from Merck, Sharpe & Dohme. D.C.W. is an investigator of the Howard Hughes Medical Institute.

The strength of galactic magnetic fields

X. Chi & A. W. Wolfendale

Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK

THE magnitudes of galactic magnetic fields are usually estimated from measurements of the radio synchrotron emission arising from acceleration of cosmic-ray electrons in the magnetic field. To interpret the emission spectrum, it is usually assumed that the energy density in the magnetic field is equal to that in the cosmic-ray protons (which are assumed to outnumber the electrons by 100:1, as they do in our Galaxy). There is, however, no compelling justification for this assumption of energy equipartition. Here we use measurements by the Compton Gamma Ray Observatory^{1,2} of the γ -ray flux from the Magellanic clouds together with radio continuum data to estimate the strength of the magnetic fields in these galaxies without having to invoke energy equipartition. We find that the assumption of energy equipartition is not valid in these irregular galaxies, and that the validity cannot be restored by changing the electron/proton ratio. Supporting evidence for our conclusion comes from radio and X-ray observations (M. G. Watson, personal communication) of the starburst galaxy M82. Our results imply that these galactic fields are too large to have been generated by dynamo action alone, and we suggest that recent star formation might instead provide the generating mechanism.

The possibility of resolving the question of energy equipartition arises from the recent measurement of γ -ray fluxes from the Magellanic clouds which are due, in part, to electron (e) and proton (p) interactions with gas in these galaxies. This allows us to estimate the cosmic-ray intensities in these galaxies independently. (A single power law is assumed to cover the whole energy range from 100 MeV to GeV for the electrons in question. The effect of solar modulation is a problem for measurement of the electron spectrum below 1 GeV, but the results from demodulation analyses and indirect measurements show that the spectrum does not seriously deviate from a power law.) We demonstrated in earlier work³ that the average cosmic-ray intensity in the Large Magellanic Cloud (LMC) is $(15 \pm 5)\%$ of the local galactic intensity, and in the Small Magellanic Cloud (SMC) it is $\leq 11\%$, and these values will be used in what follows.

The radio continuum data for the Magellanic clouds can now be examined using the above intensities of cosmic rays. If the cosmic-ray electron spectrum has a power-law form

$$N(E) dE = KE^{-\gamma} dE \quad (1)$$

where $N(E)$ is the number density in units of electrons per erg per cm^3 , then, for the case where the electrons are homogeneous and isotropic and the magnetic fields are randomly directed (but of constant mean value throughout), the intensity of synchrotron radiation is given by⁴

$$I_\nu = 1.35 \times 10^{-22} a(\gamma) LKH^{(\gamma+1)/2} \times \left(\frac{6.26 \times 10^{18}}{\nu} \right)^{(\gamma-1)/2} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1} \text{ Hz}^{-1} \quad (2)$$

where ν is the frequency of the synchrotron radiation, $a(\gamma)$ is a slowly varying function of γ and is roughly 0.085 for the range $\gamma = 2.4$ – 2.6 , L is the length of emission region along the line of sight in centimetres and H is the strength of magnetic field in gauss.

The data from a recent survey⁵ at 2.45 GHz are used here (the electrons responsible have energies in the gigaelectronvolt region). After subtracting the thermal component, the average intensity of the nonthermal emission is $1.9 \times 10^4 \text{ Jy sr}^{-1}$ over the LMC and $5.0 \times 10^3 \text{ Jy sr}^{-1}$ over the SMC. These fluxes no doubt contain a contribution from radio sources in the galaxies and

in the background, but the fraction seems to be small⁶. In choosing the geometrical parameter L , we use information from a recent review⁷ on the structure of the clouds and take $L = 2 \text{ kpc}$ (including an assumed inclination of 45°) for the LMC and $L = 10 \text{ kpc}$ for the SMC. Because the cosmic-ray electrons have a flat spectrum ($\gamma = 2.4$) in the LMC but a steep spectrum ($\gamma = 2.6$) in the SMC, different K values are used: $K = 5.12 \times 10^{-16} \text{ electrons erg}^{-1} \text{ cm}^{-3}$ for the LMC and $K = 1.13 \times 10^{-16} \text{ electrons erg}^{-1} \text{ cm}^{-3}$ for the SMC, the normalization being made at 0.36 GeV. The ratio of cosmic-ray electrons to protons above 1 GeV is taken to be the commonly used value, 1/100. Substituting these values into equation (2), we obtain the average strength of the magnetic fields in the galaxies, $H = 18.4 \mu\text{G}$ for the LMC and $H \geq 5.5 \mu\text{G}$ for the SMC. In comparison, the average magnetic field in our Galaxy, a normal spiral galaxy, is only 3–6 μG .

The field strengths correspond to an energy density two orders of magnitude higher than that of cosmic rays in the LMC, and one order of magnitude higher in the SMC. The assumption of energy equipartition is clearly violated. The problem cannot be circumvented by changing the e/p ratio, as the electrons produce γ -rays rather efficiently as well. The situation is illustrated in Fig. 1 where we present both γ -ray and radio results on the same graph. Starting with γ -rays, the curved lines give the relationship between the electron intensity and proton intensity (each in terms of the local galactic value) needed to give the measured diffuse γ -ray flux from the LMC and the upper limit to this flux measured for the SMC. Turning to the radio component, the diagonal lines represent the combination of electron flux and magnetic field required to give the measured radio flux, it being implicitly assumed that there is equipartition between the magnetic energy density and that in protons. It is evident that no e/p ratio will give agreement, by which we mean an intersection of the full or dashed lines.

In the same spirit as the above analysis, the magnetic field in the nuclear halo of the starburst galaxy M82 can be determined. This galaxy has the bonus feature that its far-infrared (FIR) emission dominates the radiation spectrum. Cosmic-ray electrons of $\sim 100 \text{ MeV}$ scatter off the FIR photons to produce

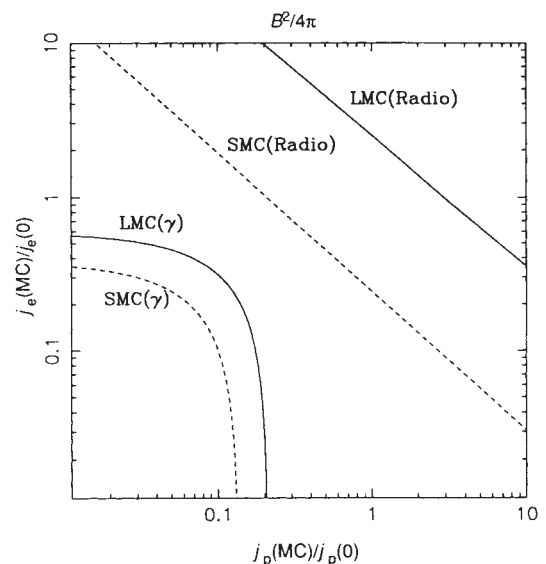


FIG. 1 Relationship between average electron intensity j_e , proton intensity j_p (both in terms of the local galactic values) and magnetic field for the Large and Small Magellanic Clouds. The magnetic field scale corresponds to the proton intensity via energy equipartition: that is, using the bottom scale, unity corresponds to $B^2/8\pi = 0.5 \text{ eV cm}^{-3}$, or $B^2/4\pi = 1 \text{ eV cm}^{-3}$. The fact that neither the full lines nor the dashed lines cross indicates that energy equipartition does not occur in either galaxy. The lines fit the radio data and the γ -ray data, as indicated.

kiloelectronvolt X-rays through inverse Compton scattering. The same electrons also interact with the galactic magnetic field to produce radio synchrotron radiation in the gigahertz region. X-ray observations by Rosat (M. G. Watson, personal communication) shows that the best-fit spectrum in the energy range 0.01–2.0 keV is that of thermal bremsstrahlung with temperature $T = 0.2$ keV and thermal electron density $n = 0.35$ electrons cm^{-3} . The corresponding X-ray emissivity in the range 1–2 keV can be used to set an upper limit on the intensity of cosmic-ray electrons.

The energy density of FIR photons in the halo of M82 is found to be 200 eV cm^{-3} using the luminosity measured by the infrared astronomy satellite IRAS⁸. The formulations of the non-relativistic bremsstrahlung and inverse Compton scattering in the Thomson limit⁹ lead to an upper limit on the cosmic-ray electron energy density $\epsilon_e = 0.006 \text{ eV cm}^{-3}$. If a power-law spectrum with index $\gamma = -2.4$ is assumed for the electrons, the spectral coefficient will be $K = 1.1 \times 10^{-14} \text{ erg}^{-\gamma-1} \text{ s}^{-1} \text{ cm}^{-3}$, which is 22 times the local galactic value. Klein *et al.*¹⁰ gave an equipartition magnetic field of $50 \mu\text{G}$, but if the above-derived cosmic-ray electron intensity is used, the magnetic field needs to be as high as $\sim 120 \mu\text{G}$ to account for the radio emission. Again, the energy equipartition assumption is violated; the field energy density is about two orders of magnitude higher than cosmic-ray energy density if the e/p ratio is assumed to be 1/100.

Three main hypotheses have been put forward so far for the origin of galactic magnetic fields. One is the dynamo mechanism¹¹, in which a weak seed field can be amplified by dynamo action in the galactic disk associated with differential rotation. Next is the idea of primordial origin¹² in which the initial (universal) field configuration is 'wound up' by the galactic rotation.

The third model involves discrete objects. Hoyle¹³ has suggested that the magnetic fields generated near a central compact object could be responsible for building up a galactic magnetic field. Daly and Loeb¹⁴ developed this idea by using jets to distribute the magnetic fluxes from a central compact sources. Rees¹⁵ proposed that the galactic magnetic field stems from the very first stars formed in the galaxy in question.

We can now examine the question of which model of field generation is appropriate. It is relevant to look at the configuration of the magnetic fields in the Magellanic clouds. Radio polarization measurements⁵ show that only a small part of the LMC is seen in the linear polarization map (implying that the regular magnetic field is small) and no definite polarization has yet been detected in the SMC. According to the dynamo amplification mechanism, the large-scale field lines should be mainly parallel to the galactic disk and the effect of the consequent large regular field would be shown in the radio polarization map. Amplification by the dynamo mechanism therefore seems unlikely to be strong for the LMC. Furthermore, there is no firm evidence⁷ for the existence of galactic rotation in the SMC (a sure requirement for the dynamo action) increasing the likelihood that the dynamo mechanism is not in action here, either. The model of primordial origin suffers from the same drawback—no 'wound up' configuration is shown on the radio polarization map.

As star formation rates have been high in both the LMC and SMC¹⁶, it is not unreasonable to postulate that the magnetic fields in these galaxies are instead generated by the star formation process. As Rees¹⁵ has pointed out, magnetic flux leakage could occur during protostellar collapse via bipolar flow. Neutron stars could spin off magnetic field by means of a relativistic wind, and this magnetic flux could be significant.

Our model involving a stellar origin can also explain the contrast of magnetic field between the spiral arm region and the inter-arm region observed in some spiral galaxies^{17,18}—the large-scale magnetic field in the spiral arm is more random than that in the inter-arm. The field fluxes in the spiral arm are largely primary, being produced by individual stellar objects, and thus

have random orientations, whereas the field fluxes in the inter-arm are mainly secondary, having been regularized from the initial field by galactic rotation, and thus show more coherence. Buoyancy and turbulent diffusion will cause the field fluxes produced in the galactic disk to rise into the halo, where field line reconnection presumably proceeds rather rapidly, thereby making the field smoother and more regular.

New observations of γ -rays from other galaxies should enable a check to be made on the present hypotheses that there is no general equipartition between the energy densities of magnetic field and cosmic rays, that galactic magnetic fields might not often be produced by dynamo action, and that the e/p ratio is not necessarily a fundamental constant. $\square$

Received 11 January; accepted 10 March 1993.

1. Sreekumar, P. *et al.* *Astrophys. J.* **400**, L67–L70 (1992).
2. Sreekumar, P. *et al.* *Phys. Rev. Lett.* **70**, 127–129 (1993).
3. Chi, X. & Wolfendale, A. W. *J. Phys. G.* (in the press).
4. Ginzburg, V. L. & Syrovatskii *The Origin of Cosmic Rays* (Pergamon, Oxford, 1964).
5. Haynes, R. F. *et al.* *Astr. Astrophys.* **252**, 475–486 (1991).
6. Ye, T. & Turtle, A. *J. Mon. Not. R. astr. Soc.* **249**, 693–697 (1991).
7. Westerlund, B. E. *Astr. Astrophys. Rev.* **2**, 29–78 (1990).
8. Rice, W. *et al.* *Astrophys. J. Suppl.* **68**, 91–128 (1988).
9. Heitler, W. *The Quantum Theory of Radiation*, 3rd edn (Clarendon, Oxford, 1954).
10. Klein, U., Wiebeleski, R. & Morsi, H. W. *Astr. Astrophys.* **190**, 41–46 (1988).
11. Parker, E. N. *Astrophys. J.* **163**, 255–278 (1971).
12. Paddington, J. H. *Mon. Not. R. astr. Soc.* **128**, 345–360 (1964).
13. Hoyle, F. *Nature* **223**, 936–936 (1969).
14. Daly, R. A. & Loeb, A. *Astrophys. J.* **364**, 451–455 (1990).
15. Rees, M. J. *Q. J. R. astr. Soc.* **28**, 197–206 (1987).
16. Lequeux, J. in *Structure and Evolution of the Magellanic Clouds*, IAU Symp. 108 (eds van den Bergh S. & de Boer, K. S.) 67–78 (Reidel, Dordrecht, 1984).
17. Chi, X. & Wolfendale, A. W. *J. Phys. G*, **16**, 1409–1421 (1990).
18. Sukumar, S. & Allen, R. J. *Nature* **340**, 537–539 (1989).

ACKNOWLEDGEMENTS. We thank C. Fichtel for providing preprints prior to publication, M. Watson for correspondence on M82 and J. Osborne for helpful discussions.

Structural distortion in Rb_3C_{60} revealed by ^{87}Rb NMR

R. E. Walstedt, D. W. Murphy & M. Rosseinsky*

AT&T Bell Laboratories, Research Division, 600 Mountain Avenue, Murray Hill, New Jersey 07974-0636, USA

INTEREST in the alkali-intercalated A_xC_{60} fullerides (where A is an alkali metal and $x = 2, 3, 4, 6$)^{1,2} has been greatly enhanced by the discovery of a family of superconducting phases for $x = 3$ with transition temperatures (T_c s) that are exceeded only by the copper oxide materials^{3–9}. An early structure refinement⁵ identified K_3C_{60} as a face-centred-cubic (f.c.c.) intercalation phase with two tetrahedral (T) and one octahedral (O) potassium ion per C_{60} . The other superconducting A_3C_{60} systems are isostructural⁷. Here we report ^{87}Rb NMR studies of Rb_3C_{60} which show the expected T:O ratio of 2:1 at 440 K. Below 370 K, however, an additional resonance appears; spin-echo double-resonance NMR experiments^{11,12} show that this arises from a second type of tetrahedral Rb^+ site (T'). Thus the bulk of this material has a local symmetry lower than that of the f.c.c. structure derived from X-ray refinements. Models that could account for this include a rotational orientation of C_{60} that places a C_5 ring adjacent to the T' Rb^+ ions. Alternatively, the symmetry lowering might be due to a Jahn–Teller distortion on C_{60}^{3-} , carrier-density modulation or Rb^+ clustering. Each of these models could lead to a degree of charge localization and consequent reduction in the density of states and hence in T_c .

Samples of Rb_3C_{60} and one of $\text{Rb}_{1.5}\text{K}_{1.5}\text{C}_{60}$ were prepared by direct reaction of A_6C_{60} and C_{60} at 250–350 °C for 10–15

* Present address: Inorganic Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, UK.

days. X-ray powder diffraction studies at room temperature verified the bulk f.c.c. phase composition. A 40% diamagnetic shielding fraction with onset at $T \approx 29.5$ K (Rb_3C_{60}) also indicated good sample quality. Phase purity is estimated to be at least 95%. One of these samples was used in earlier ^{13}C nuclear magnetic resonance (NMR) studies¹⁰.

^{87}Rb NMR spectra taken at $T = 200$ K are shown in Fig. 1 for specimens of Rb_3C_{60} (a) and $\text{Rb}_{1.5}\text{K}_{1.5}\text{C}_{60}$ (b). The inset to Fig. 1a shows the ^{87}Rb NMR spectrum at $T = 443$ K, where we see that it consists of only the T and O lines, in the stoichiometric intensity ratio of 2. Here the ^{87}Rb lines are narrowed by rapid rotation of the C_{60} molecules. As the temperature is lowered the T' line gradually appears. At $T = 200$ K the spectrum includes the O line centred at a shift $K \approx 40$ p.p.m., the T line at $K \approx 165$ p.p.m., and, in addition, the T' line at $K \approx 270$ p.p.m. These three features broaden gradually with lowering temperature, but retain fixed relative positions and intensities in a ratio of 35:55:10 (each ± 2) over the range $150 \text{ K} \leq T \leq 300 \text{ K}$. The chief source of uncertainty in these weights stems from partitioning the small overlap between the lines. We note in passing that below T_c the ^{87}Rb lines broaden further and move rapidly to lower shift values in a fashion similar to the ^{13}C resonance¹⁰. The relative intensities of the O and T lines in the mixed (rubidium, potassium) sample (Fig. 1b) are markedly different, evidently modified by a preferential occupation of the T sites by K^+ ions on account of their smaller size. The T' line for this sample is broader, but has roughly the same intensity relative to the T line as in Fig. 1a. The relative intensity of the T' line has been reproduced with two independently prepared samples of Rb_3C_{60} and is, we conclude, an intrinsic feature. By observing a NMR spectrum identical to that in Fig. 1a at a frequency 20% lower, we have verified that quadrupolar splitting and broadening effects are negligible.

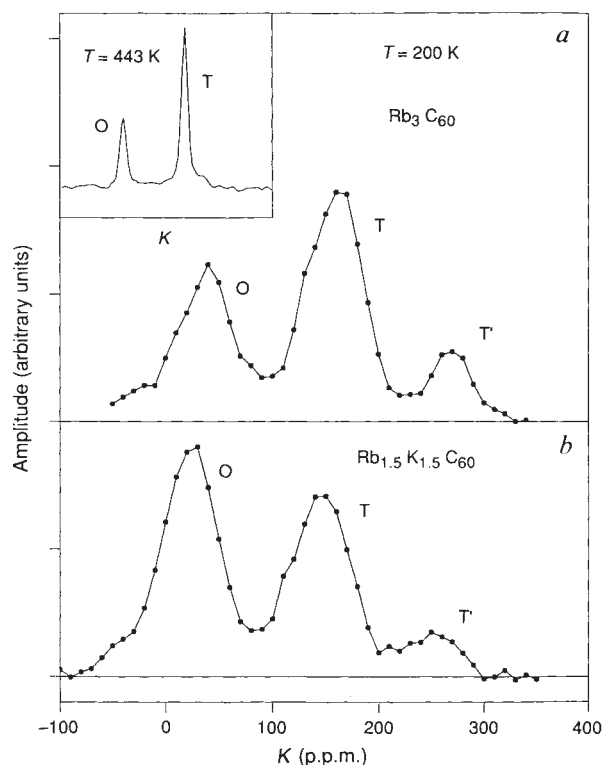


FIG. 1 ^{87}Rb NMR spectra for powder samples of Rb_3C_{60} (a) and $\text{Rb}_{1.5}\text{K}_{1.5}\text{C}_{60}$ (b) at $T = 200$ K and a NMR frequency of 101 MHz. Two independently synthesized samples of Rb_3C_{60} gave indistinguishable results. Shifts are referred to the gyromagnetic ratio of the $^{87}\text{Rb}^+$ ion in the gas phase, $\gamma/2\pi = 1.392812 \text{ kHz gauss}^{-1}$. The inset shows the ^{87}Rb NMR spectrum in Rb_3C_{60} at $T = 443$ K, where the O and T lines are motionally narrowed, and the T' line is no longer visible.

TABLE 1 Signal decrements for the Rb_3C_{60} spectrum

Observe	Experimental data		Estimated decrements		
	Pulse	SEDOR data	(1)	(2)	(3)
O	T	0.20 ± 0.02	0.27	0.24	0.27
T	O	0.10 ± 0.01	0.14	0.14	0.09
T'	T	0.08 ± 0.02	—	0.10	0.24
T'	O	0.095 ± 0.02	—	0.14	0.014

Measured signal decrements for all three peaks in the Rb_3C_{60} spectrum are tabulated for SEDOR pulses applied at the O and T peak frequencies. For comparison, calculated estimates for three cases described in the text are also listed: (1) the unmodified f.c.c. phase with no T' site; (2) the T' site as a tetrahedral site; and (3) the T' site as an octahedral site. Option 2 is clearly the only one compatible with the experimental data.

To examine the spatial relationship of the O, T and T' sites in the Rb_3C_{60} lattice, we have carried out spin-echo double-resonance (SEDOR) experiments^{11,12} at $T = 200$ K. In this technique, a spin echo is observed by selectively exciting one of the three resonance lines in Fig. 1, which we shall call the A spins. An additional excitation pulse is applied during the echo sequence which selectively inverts spins in one of the other lines ('B' spins), without exciting transitions in the other two lines. If the A and B spins are close neighbours in the lattice, and thus have appreciable dipolar coupling between them, then the third (SEDOR) excitation pulse will disrupt the refocusing of the echo. As a result, one observes a decrement in the echo amplitude roughly proportional to z_B/r_{AB}^6 , where r_{AB} is the distance between an A spin and its nearest-neighbour B spins, and z_B is the number of B nearest neighbours. The SEDOR method is therefore a quantitative tool for determining the distance between sites corresponding to different NMR lines in the spectrum.

Measured values are given in Table 1 for the decrements of echoes from all three lines shown in Fig. 1 when SEDOR pulses are applied to the O and T lines. The measured value for the largest decrement, which is the one found when observing the O line and pulsing the T line, is slightly smaller than, but in reasonable agreement with a calculated estimate given in columns B and C. From the data given, we can conclude immediately that all three sites are intermingled on an atomic scale. Thus, there is no question of different chemical phases contributing to the spectra of Fig. 1. The T' site must therefore arise from special orientations of the C_{60} molecules or from some other deviation from the ideal f.c.c. structure of Rb_3C_{60} .

To evaluate the data further, calculated SEDOR decrements (scaled to the largest one) for three hypothetical structures are also listed in Table 1. These are: (1) the case of no disorder, where the T' line is necessarily absent; (2) the case where the T' line arises from a modified T site; and (3) the case where the T' line arises from a modified O site. In cases (2) and (3) the nature of the modification is as yet unspecified and the interionic distances for stoichiometric Rb_3C_{60} are assumed to hold. In case (1) the O and T decrements differ by a factor of ~ 2 , in rough agreement with the data. This verifies to zero order the identities of the two main lines. Comparing cases (2) and (3), one finds very strong evidence that the T' line is a modified T site: the predicted values for an O site are well outside the error limits, whereas those for a T site are in good agreement. This conclusion is also supported by the observed intensities mentioned earlier. Summing the intensities of the T' and T lines, one finds a ratio of 65:35 for total tetrahedral to octahedral site occupation. This is within experimental error of the 2:1 ratio expected for stoichiometric Rb_3C_{60} .

Further experiments on the dynamic NMR behaviour of the lines in Fig. 1 give added insight into the nature of the T' sites. In a spin-lattice relaxation measurement, the polarization of

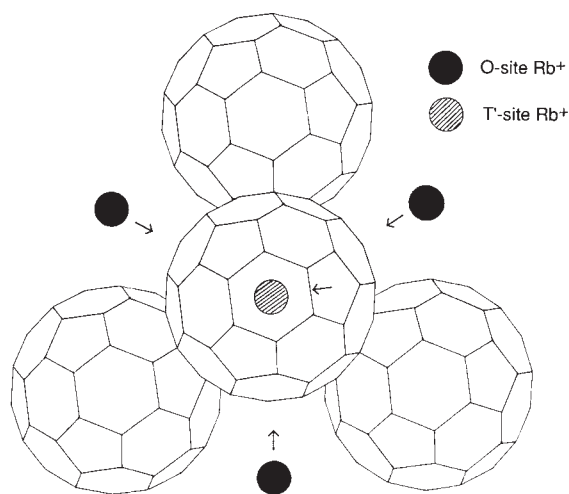


FIG. 2 View of the molecular complex surrounding tetrahedral sites from along a (111) axis. The C_{60} molecules are shown oriented with hexagonal faces normal to the four (111)-equivalent directions. The T' site at the centre is viewed through hexagonal faces of the C_{60} in the foreground. The possible clustering of neighbouring O-site Rb^+ ions is indicated by arrows. The arrow on the C_{60} in the foreground shows a rotation which would place C_5 rings along the (111) axis.

recovery of a resonance line is monitored following the application of a saturation pulse. In the closely related saturation transfer experiment, one monitors the migration of nuclear spins from a line that has been saturated to a different line in the spectrum. This transfer takes place with a time constant characteristic of the motion of spins between sites in the lattice. In the present instance, we observe saturation transfer between the T and T' lines starting close to room temperature with a rate which increases rapidly at higher temperatures. The temperature variation is consistent with thermally activated motion with an energy barrier of $\Delta \approx 0.5$ eV. There is no measurable transfer between the O site and the other two, so migration must occur only between the T and T' sites. This rules out chemical diffusion between the latter sites, because any such process would necessarily involve the O sites that surround the T and T' sites. The saturation transfer effects are therefore a result of T' site migration in the lattice. If, for example, the T and T' sites are distinguished by the orientation of their C_{60} neighbours, then the T' sites could migrate by a process of coordinated reorientation of the C_{60} molecules. This would result in an exchange of nuclear spins between the T and T' lines without involving the O-site nuclei, as is observed.

Similar experiments have also revealed spectral diffusion of the NMR frequencies within the O and T line profiles. This effect persists to well below room temperature and is physically distinct from the saturation transfer process. The spectral diffusion is probably due to thermally activated reorientation of the C_{60} molecules. Thus, the varying C_{60} coordination geometries presented to the Rb^+ ions in the O and T sites generate a distribution of local NMR shifts (line broadening). This, we suggest, is the primary source of broadening for the T and O line. It follows that the diminished NMR linewidth observed for the T' line is good evidence for a unique C_{60} coordination at those sites.

The structural model for A_3C_{60} (where A is K or other alkali metal)⁵ includes merohedral disorder of two C_{60} orientations,

each of which presents C_6 rings of the C_{60} to the TRb^+ ions, making the local coordination $A[C_6]_4$ as shown in Fig. 2. This model, with rapid reorientation of the C_{60} molecules is consistent with the ^{87}Rb NMR spectrum at elevated temperatures (Fig. 1a, inset). But the appearance of the (T') site (Fig. 1) requires a substantial local structural change. This effect may contribute to large thermal parameters and diffuse scattering observed in X-ray⁷ and neutron diffraction¹³ studies of polycrystalline samples. Possible origins for the structural changes include displacements of the Rb^+ or C_{60} ions, symmetry lowering of C_{60}^{3-} resulting from a Jahn-Teller distortion, disproportionation¹⁴ of C_{60}^{3-} [$2C_{60}^{3-} \rightleftharpoons C_{60}^{2-} + C_{60}^{4-}$], or a change in the local $Rb(C_6)_4$ coordination. The ^{13}C NMR¹⁰ seems to rule out a large Jahn-Teller effect on C_{60}^{3-} , however.

A few main ideas emerge from the range of possibilities. First, displacement of all four O Rb^+ neighbours towards a TRb^+ would create two types of T site in a 7:1 ratio, close to the observed value. A second possibility is orientation of the C_{60} molecules so that C_5 rings are presented to (a maximum of) 25% of the TRb^+ , giving the remaining TRb^+ a distribution of coordination geometries. This would account for the fact that the T' line is only about half the width of the T line. The observed 5.5:1 ratio for T:T' sites is consistent with the C_5 ring hypothesis if the orientation effect is less than 100%.

The increased NMR shift at the T' sites is presumably a ^{87}Rb contact shift from carrier orbitals on neighbouring C_{60} molecules. As such, it is not clear how the strong T-T' shift contrast could arise without an increased carrier density near the T' sites. Put another way, the T-T' splitting is what one would expect to observe if a charge density modulation were present¹⁵. The presence of any such effect would undoubtedly modify the density of states at the Fermi surface and, thus, the superconducting transition temperatures in these materials. Any theory of superconductivity in these materials will therefore have to take into account this lowering of symmetry and its effect on the band structure at the Fermi surface. □

Received 30 November 1992; accepted 16 March 1993.

- Fullerenes, ACS Symp. Ser. 481 (eds Hammond, G. S. & Kuck, V. J.) (Am. chem. Soc., Washington DC, 1992).
- Murphy, D. W. *et al.* *J. Phys. Chem. Solids* **53**, 1321-1332 (1992).
- Hebard, A. F. *et al.* *Nature* **350**, 600-601 (1991).
- Holczer, K. *et al.* *Science* **252**, 1154-1157 (1991).
- Stephens, P. W. *et al.* *Nature* **351**, 632-634 (1991).
- Rosseinsky, M. J. *et al.* *Phys. Rev. Lett.* **66**, 2830-2832 (1991).
- Fleming, R. M. *et al.* *Nature* **352**, 787-788 (1991).
- Tanigaki, K. *et al.* *Nature* **352**, 222-223 (1991).
- Kelty, S. P., Chen, C.-C. & Lieber, C. M. *Nature* **352**, 223-224 (1991).

- Tycko, R. *et al.* *Phys. Rev. Lett.* **67**, 1886-1889 (1991).
- Kaplan, D. & Hahn, E. L. *J. Phys. Radium* **19**, 821-826 (1958).
- Walstedt, R. E. & Wernick, J. H. *Phys. Rev. Lett.* **20**, 856-859 (1968).
- Christides, C. *et al.* *Phys. Rev. B* **46**, 12088-12091 (1992).
- Zhang, F. C., Ogata, M. & Rice, T. M. *Phys. Rev. Lett.* **67**, 3452-3455 (1991).
- Pfeiffer, L., Walstedt, R. E., Bell, R. F. & Kovacs, T. *Phys. Rev. Lett.* **49**, 1162-1165 (1982).

ACKNOWLEDGEMENTS. We thank S. E. Barrett and R. Tycko for discussions and for doing the high-temperature ^{87}Rb measurements reported here. We also thank P. Littlewood, A. Pasquarello, T. M. Rice and J. Zaanen for discussions.

Separation of chiral phases in monolayer crystals of racemic amphiphiles

C. J. Eckhardt*, N. M. Peachey*, D. R. Swanson*,
J. M. Takacs*, M. A. Khan*, X. Gong*, J.-H. Kim‡,
J. Wang‡ & R. A. Uphaus‡

* Department of Chemistry, University of Nebraska-Lincoln, Lincoln, Nebraska 68588, USA

‡ Ames Laboratory, Iowa State University, Ames, Iowa 50011, USA

OPTICAL activity manifested by chiral molecules and crystals is of long-standing interest in physics, chemistry, biology and geology^{1,2}. The structure of chiral lattices in two and three dimensions may provide insights into chiral discrimination, and chiral phases play an important part in the physics and applications of liquid-crystal and amphiphilic films³⁻⁷. Here we report the observation of spontaneous separation of chiral phases in an oriented monolayer of rigid, chiral amphiphiles deposited on mica. Atomic force microscopy of the ordered films reveals domains of mirror-image structures. We propose that this may be the signature of separation into domains of pure enantiomers, although other possibilities exist. This separation of chiral phases in two dimensions may be considered analogous to the spontaneous resolution of enantiomers in three-dimensional crystallization, as exploited by Pasteur in 1848 to isolate enantiomers of sodium tartrate¹.

Seminal studies⁴ of monolayers of chiral fatty acid derivatives on aqueous subphases demonstrated both that the monolayers showed crystallinity and that racemates and pure enantiomers formed different solid phases, but experimental limitations prevented further characterization. Atomic force microscopy (AFM)⁶ later showed that pure enantiomer phases had greater surface viscosity and long-range order, but molecular resolution was not achieved.

One of our aims is to create amphiphiles with which to exploit symmetry constraints of the 17 two-dimensional space groups for the deliberate formation of specific planar lattices. For chiral asymmetric units, the oblique and rectangular classes (Fig. 1)

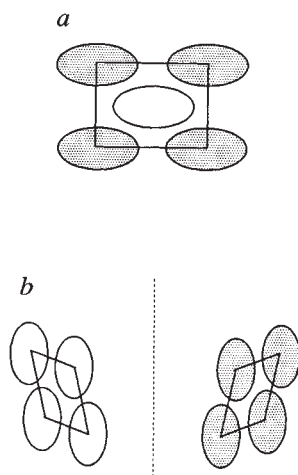


FIG. 1 Schematic drawings of (a) a two-molecule rectangular unit cell; (b) the mirror-image relationship of antipodal oblique unit cells. The mirror line is indicated by dashes. The two oblique cells cannot be superimposed by any two-dimensional symmetry operation. Antipodal units are indicated by different cross-hatching in all of the drawings.

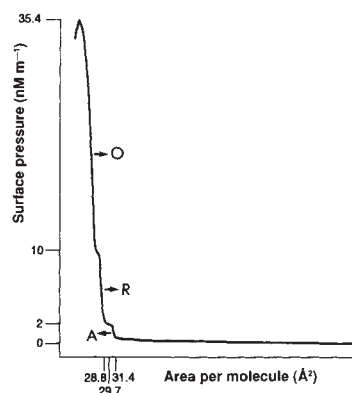
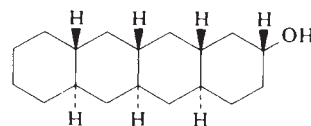


FIG. 2 Surface pressure isotherm for a TCA monolayer formed at 25 °C on a conventional Langmuir trough in a clean room. The subphase is Milli-Q water. The A_{aq} -, R_{aq} - and O_{aq} -phases are labelled and are delimited by narrow plateaus.

have mutually exclusive requirements⁸. The former can pack only pure enantiomers whereas the latter only accommodates racemates. These space-group constraints are retained even if the asymmetric unit includes more than one three-dimensionally chiral molecule. We synthesized a chiral tetracyclic alcohol (TCA) of elliptical cross-section as a racemate by an iterative Robinson annulation starting from 2-decalone:



TCA has proved resistant to optical resolution by classical methods.

The TCA surface isotherm (Fig. 2) indicates the presence of three phases which we designate as A_{aq} , R_{aq} and O_{aq} (amorphous, rectangular and oblique phases respectively; subscript aq denotes a phase on the water surface) with areas per molecule of 31.4, 29.8 and 28.4 Å², respectively. The collapse pressure indicates a strength similar to that of fatty acid films.

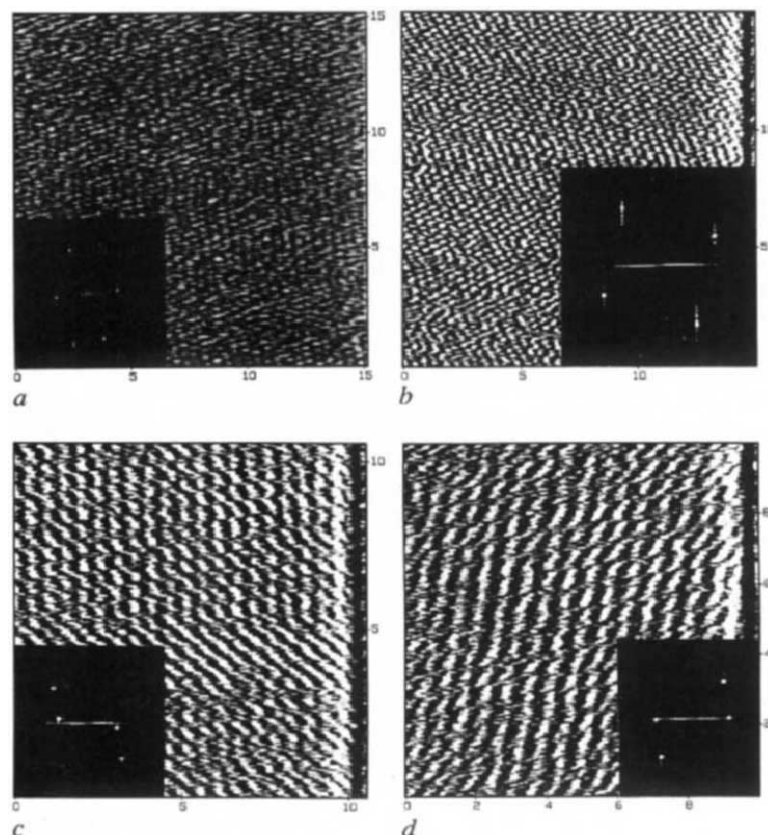
We transferred monolayers of each of the Langmuir film phases onto mica. This process may cause a change in packing⁹ and also be substrate-specific¹⁰, so we refer to the corresponding Langmuir-Blodgett monolayer films on the mica substrate as A_m , R_m and O_m .

The unenhanced image (Fig. 3a) of a 15 × 15 nm area of the A_m -phase is disordered with a correspondingly poor Fourier pattern. Regions of considerable disorder appear in which are embedded more-organized, apparently rectangular, sections. There are many domains and the entire film has a glass-like character.

The unenhanced image of the R_m -phase film, again taken over a 15 × 15 nm area, is more ordered as evidenced by its Fourier pattern (Fig. 3b) which shows centred rectangular packing (Fig. 1a). The constants for a unit cell of two molecules are $a = 10.0 \pm 0.5$ Å, $b = 6.1 \pm 0.4$ Å and $\gamma = 91^\circ \pm 4^\circ$ with an apparent planar space group of pg . The area per molecule is 30.5 ± 8.2 Å². This lattice is incommensurate with the mica cleavage plane lattice parameters of $a = 5.194$ Å and $b = 8.996$ Å.

The unenhanced image O_m -phase film for a 10 × 10 area yields the best Fourier pattern (Fig. 3c). The averaged values for the experimental lattice constants for this section are: $a = 5.52 \pm 0.32$ Å, $b = 7.71 \pm 0.46$ Å and $\gamma = 41^\circ \pm 3^\circ$. The measured area per molecule is 27.9 ± 20 Å². Such errors are typical of dimensions determined by AFM^{11,12}. The O_m -phase image of a larger area (Fig. 4) shows evident grain boundaries between

FIG. 3 AFM images with corresponding Fourier pattern inserts of TCA monolayer films vertically transferred at 3 mm min^{-1} onto freshly cleaved, atomically flat, mica (muscovite 2M₂ polymorph) plates. Images were obtained using a Nanoscope II AFM with selected silicon nitride scanning tips of $100 \text{ \AA}$ radius of curvature which were mounted on a cantilever of 0.38 N m^{-1} force constant. The AFM was used in constant height mode with forces ranging from 10 to 50 nN. Greater tip forces caused disruption of the film but at no time was the amphiphile scraped from the surface. The instrument was calibrated with mica (muscovite 3T polymorph) which was scanned 40 times, as were the films, with subsequent determination of average lattice values. Thermal drift and tip drag were characterized and compensated for in the measurements. AFM image and Fourier pattern of (a) a $15 \times 15 \text{ nm}$ area of the monolayer A_m-phase transferred to mica at 1.8 mN m^{-1} ; (b) a $15 \times 15 \text{ nm}$ area of the monolayer R_m-phase transferred to mica at 7.5 mN m^{-1} ; (c) a $10 \times 10 \text{ nm}$ area of the monolayer O_m-phase transferred to mica at 15.4 mN m^{-1} ; (d) a $10 \times 10 \text{ nm}$ area of the same monolayer as in c but from a region where the packing is a mirror-image of that observed in c. The Fourier patterns in c and d differ in general shape because of relative orientation of the regions to the scan direction of the AFM tip. Patterns for the O_m-phase are the best; those for the A_m-phase are the worst. The larger errors in the R_m-phase are due to shorter orientational and translational coherence lengths of the film.



crystalline regions that differ in the senses of their cells. Another $10 \times 10 \text{ nm}$ area of the film (Fig. 3d) has a different sense of packing from the previous one. The cells have the same dimensions but are mirror images of those in the other area.

Net symmetries and symmetry requirements imposed by chiral systems afford straightforward interpretation of these results. The A_m-phase is a disordered solid phase of TCA. On increasing surface pressure, the molecules transfer as monolayer, rectangular crystals. Because rectangular nets cannot accommodate asymmetric units of pure enantiomers, the R_m-phase must be that of the racemate. The O_m-phase, which is transferred at the highest surface pressure, forms intergrowths of mirror-image oblique planar lattices. Such chiral intergrowths are known in three dimensions¹³. The existence of two mirror-image crystals (Fig. 3c, d; 1b) in the O_m film is direct proof of separate chiral phases.

Theory shows³ that chiral phase separation can occur for a monolayer if it is (1) a hexatic phase with tilt direction between 0° and 30° from a local bond direction, (2) non-hexatic but with inequivalent, mirror-image, molecular packings on the surface, or (3) formed of chiral domains each containing just one type of enantiomer. The last is a particular case of (2) which is itself a requirement of two-dimensional space-group symmetry. Case (1) arises because of the finite thickness of any real film.

It is difficult to assess whether the phase separation occurs during compression of the Langmuir film or on transferral to mica. To aid in this, calculations of the atom-atom potential for organic monolayers¹⁴ with no substrate interactions have been made for both racemate and enantiomer. The rigidity of TCA permits neglect of intramolecular degrees of freedom. Molecules are related by two-dimensional translations and may freely rotate about the vertical inertial axis along their length and perpendicular to the substrate. Allowing the inertial axis to relax from perpendicularity yields no significant tilt. The amphiphile cross-section and chirality then determine the long-range

packing. We calculated the minimum potential energy lattice for each of the 17 planar groups using a Buckingham potential with 'universal' parameters¹⁵. Unit cells containing up to four molecules were calculated. Among these were asymmetric units comprising two antipodal enantiomers.

For the racemate, a rectangular net of *pg* symmetry is predicted. Calculated constants are $a = 9.5 \text{ \AA}$ and $b = 5.8 \text{ \AA}$ for two molecules in the unit cell with an area per molecule of $27.6 \text{ \AA}^2$ (Fig. 1a). An oblique net, *p1*, with one enantiomer in the asymmetric unit, is the lowest-energy chiral lattice, with constants $a = 5.64 \text{ \AA}$, $b = 8.13 \text{ \AA}$ and $\gamma = 36.2^\circ$. The area per molecule is $27.0 \text{ \AA}^2$. Theoretical values are smaller than experimental ones because the calculation assumes an equilibrium system at 0 K.

Calculated cell constants and areas per molecule agree remarkably well with experiment. Only the angle value of the oblique system lies slightly outside of error estimates. The most important agreement is that of the space groups. Further, the energies of the oblique and rectangular lattices differ by $\sim 7.7 \text{ kJ mol}^{-1}$ with the former being the more stable. Because both entropic and dynamic effects are neglected in the calculation, the role of increased compression in the formation of the oblique lattice is unsurprising.

The accord of the lattice constants, space group symmetries and areas per molecule for the Langmuir, Langmuir-Blodgett and calculated Langmuir films suggests that the nets for both the aqueous and solid subphases are the same. This implies that chiral phase separation occurs on the aqueous subphase and not during transfer.

From this and the presence of intergrowths, we attribute the formation of the O_{aq}-phases to chiral phase separation of the R_{aq}-phase. The absence of hexatic packing and the improbability of inducing tilt with increased surface pressure argues against possibility (1). A pair of antipodal, three-dimensionally chiral enantiomers forming the asymmetric unit is in accord with (1).

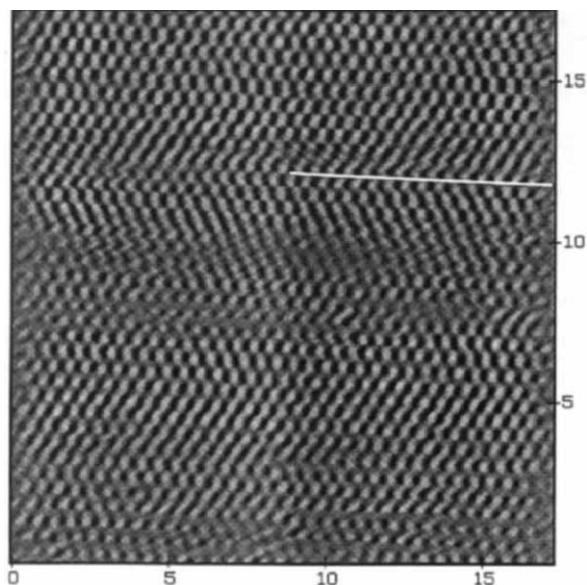


FIG. 4 Enhanced AFM image of a 17×17 nm area of the O_m -phase monolayer film deposited on mica. Grain boundaries, one of which is indicated by a line, are evident. The similarity of the banded domains to predicted stripe structures³ is striking. Enhancement will only introduce regularity and not disorder.

In two dimensions, asymmetric units composed of such a pair of three-dimensionally chiral enantiomers are achiral. This is because all enantiomers have their headgroups constrained to the same plane, and the needed mirror relationship to generate an achiral asymmetric unit can be achieved. Such a chiral pair may be generated by shear deformation of R_m . However, our atom-atom potential calculations suggest that this arrangement will not be adopted. The remaining choice, strongly supported by the calculation of the O_m -phase, is chiral separation into pure enantiomers. Decisive assignment must await diffraction experiments.

Although much work on chiral amphiphiles exists^{4,6}, we are not aware of previous direct evidence, such as given here, for spontaneous chiral resolution of type (3) in two dimensions. This may be because the amphiphiles studied previously achieve chirality only because of a chiral substituent. The intermolecular forces leading to chiral discrimination in such molecules are thus confined to one chiral centre. In contrast, TCA has an intermolecular potential which may be regarded as globally chiral for the molecule because its rigidity and polymer-like structure create a series of equivalent chiral centres. This may increase the chiral discrimination and the probability of spontaneous optical resolution. □

Received 13 October 1992; accepted 3 March 1993.

1. Pasteur, L. *C.R. Acad. Sci. Paris* **26**, 535–539 (1848).
2. Collet, A., Brienne, M.-J. & Jacques, J. *Chem. Rev.* **80**, 215–230 (1980).
3. Selinger, J. V., Wang, Z.-G., Bruinsma, R. F. & Knobler, C. M. *Phys. Rev. Lett.* **70**, 1139 (1993).
4. Lundquist, M. *Arkiv Kem.* **17**, 183–195 (1960); **21**, 395–406 (1962); **23**, 299–306 (1963).
5. Arnett, E. M., Harvey, N. G. & Rose, P. L. *Acc. chem. Res.* **22**, 131–138 (1989).
6. Knobler, C. M. & Desai, R. C. *A. Rev. phys. Chem.* **43**, 207–236 (1992).
7. Qiu, X., Ruiz-Garcia, J., Stine, K. J., Knobler, C. M. & Selinger, J. V. *Phys. Rev. Lett.* **67**, 703–708 (1991).
8. *International Tables for X-ray Crystallography*, Vol. 1 (eds Henry, N. F. M. & Lonsdale, K.) Ch. 2 (Kynoch, Birmingham, 1952).
9. Engel, M., Merle, H. J., Peterson, I. R., Riegler, H. & Steitz, H. *Ber. Bunseng. Phys. Chem.* **95**, 1514–1520 (1991).
10. Schwartz, D. K., Garnaes, J., Viswanathan, R. & Zasadsinski, J. A. *N. Science* **257**, 508–511 (1992).
11. Meyer, E. *et al. Nature* **349**, 398–400 (1991).
12. Bourdieu, L., Silberzan, P. & Chatenay, D. *Phys. Rev. Lett.* **67**, 2029–2032 (1991).
13. Ramdas, S., Thomas, J. M., Jordan, M. E. & Eckhardt, C. J. *J. phys. Chem.* **85**, 2421–2425 (1981).
14. Eckhardt, C. J. & Swanson, D. R. *Chem. Phys. Lett.* **194**, 370–374 (1992).
15. Pertsin, A. J. & Kitaigorodskii, A. I. *The Atom-Atom Potential Method* **89** (Springer, Berlin, 1987).

ACKNOWLEDGEMENTS. We thank E. Bahar and the Center for Electro-optical Studies of the University of Nebraska-Lincoln for use of the AFM. N.M.P. and D.R.S. thank the University of Nebraska Foundation for graduate fellowships. This research was supported by the Chemical and Biological Sciences Division of the Army Research Office and in part by the Office of Basic Energy Sciences, U.S. Department of Energy and University of Nebraska-Lincoln Center for Materials Research and Analysis. Computer time was provided by the Cornell National Supercomputer Facility.

Change in crystallization mechanism at the glass transition of colloidal spheres

W. van Meegen & S. M. Underwood

Department of Applied Physics, Royal Melbourne Institute of Technology, Melbourne, Victoria 3001, Australia

THE crystallization of 'hard' spheres, which interact only through repulsive forces on contact, is a purely entropic effect and provides a model system for testing theories of freezing transitions in liquids generally¹. Micrometre-sized colloidal particles can be made to approximate hard spheres by grafting of polymers to their surfaces¹. Their slow dynamics make them attractive systems for the study of crystallization kinetics^{2–4}. As well as undergoing crystallization^{1,2}, such particles also exhibit a glass transition at a well-defined volume fraction ϕ_g , which we have shown previously to be related to the cessation of large-scale diffusion on experimental timescales^{5,6}. Here we show that the glass transition coincides with a change in the mechanism of crystallization. At volume fractions less than ϕ_g isometric crystals are nucleated homogeneously throughout the sample, whereas above ϕ_g crystals forming slowly in the metastable glass phase are highly asymmetric. Regular rocking of the suspension above the glass transition induces more rapid formation of plate-like crystals, indicating that their nucleation is shear-induced.

We prepared nearly transparent concentrated suspensions, suitable for light scattering studies, by closely matching the refractive index of the suspending liquid to that of the particles^{1,5}. Thin layers of solvated polymer grafted to the particle surfaces prevent irreversible aggregation and provide short-ranged repulsive interparticle forces that, in turn allow the suspension concentrations to be accurately expressed in terms of hard-sphere volume fractions^{1,6,7}, ϕ . Thus the experimental freezing and melting concentrations (or packing fractions), $\phi_f = 0.494$ and $\phi_m = 0.543$, respectively, are in accord with those for a system of perfect hard spheres⁸.

Figure 1 shows photographs of crystallized suspensions with concentrations (ϕ) ranging from 0.535 (below melting) to 0.587. Studies of both light scattered through small angles by large-scale concentration fluctuations⁴ and the orientationally averaged main (111) Bragg reflection suggest that, for concentrations up to 0.574, crystallization proceeds by relatively rapid nucleation and growth, followed by slow ripening. The number density of nuclei increases with particle concentration, limiting subsequent growth to progressively smaller crystals. These roughly isometric crystals, evident within about one hour after tumbling, are hardly visible at $\phi = 0.574$. Remarkably, on increasing the concentration by about 1%, to $\phi = 0.581$, the crystals become large and highly asymmetric. While these crystals take more than 10 hours to develop in the undisturbed amorphous metastable phase, regular rocking has the spectacular effect of inducing needle-shaped crystals, parallel to the (vertical) cuvette walls, within minutes.

In Fig. 2 we show the autocorrelation functions, $f(\tau)$, of particle concentration fluctuations measured by dynamic light scattering on the metastable fluid samples before crystallization. There is a dramatic slowing in the relaxation of concentration fluctuations with increasing suspension concentration. For $\phi = 0.574$ $f(\tau)$ decays on a timescale of $\sim 1,000$ s. For $\phi \geq 0.581$, however, $f(\tau)$ saturates to an almost constant value at long times

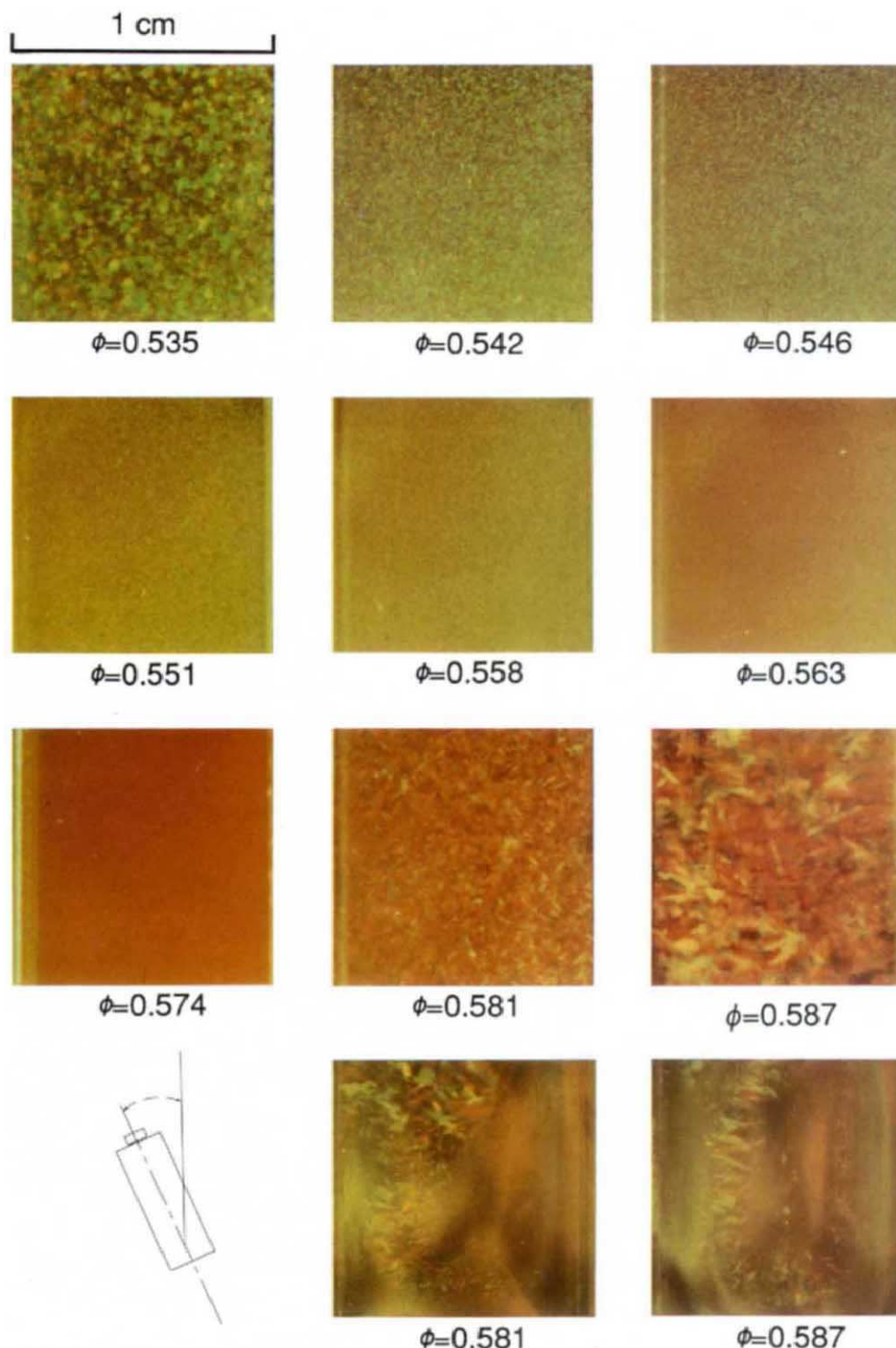
indicating that large-scale diffusion has effectively stopped for the duration of the 1,000 s experiment.

Motion in dense fluids (molecular or colloidal) can be pictured simplistically as a sequence of progressively slower processes: small-scale motion of particles confined to transient cages of neighbouring particles; collective cage motion and breakdown of the cages leading to large-scale particle diffusion^{5,6}. These three processes are not discernible at low concentrations but at higher concentrations their timescales lengthen and separate. In Fig. 2 one can identify the crossover times between these processes by points of inflection, which occur (for $\phi = 0.574$) at approximately $10^4 \mu\text{s}$ and $10^7 \mu\text{s}$. At the glass transition con-

centration ϕ_g ($0.574 < \phi_g < 0.581$) the slowest process, the breakdown of particle cages, ceases. This results in permanent caging of particles, cessation of large-scale diffusion and freezing of the fluid structure. At the same concentration crystallization by homogeneous nucleation is suppressed, implying that this nucleation proceeds by large-scale particle diffusion.

In colloidal suspensions the randomizing effect of brownian motion is easily dominated by the application of shear stresses, resulting in partial alignment of particles in layers or strings parallel to the direction of the stress⁹, with a reduction in the apparent viscosity. We conclude that the tumbling method employed to randomize the particle positions generates small,

FIG. 1 Photographs of crystallized suspensions. These contain particles ($0.2 \mu\text{m}$ radius) of poly-methyl-methacrylate suspended in a liquid mixture of decalin and carbon disulphide, chosen to provide near transparency. Sample concentrations, expressed as effective hard-sphere volume fractions ϕ , are indicated. Photographs were taken of samples left undisturbed for two days after randomization of particle positions by tumbling the samples on a wheel rotating at $\sim 1 \text{ Hz}$ in a vertical plane. Note that two photographs are shown for each of the highest two concentrations. In the last two photographs (also after two days) the samples were subjected to a small-amplitude regular rocking motion (schematically indicated in the lower left of this figure) for several minutes immediately after tumbling. The progressive change in colour of the Bragg-reflecting crystals, from green-yellow to yellow-orange in the first nine pictures, is a consequence of the optical arrangement in which the sample cells and white light source (used to illuminate the samples obliquely from behind) were fixed and the camera was moved to take the photographs.



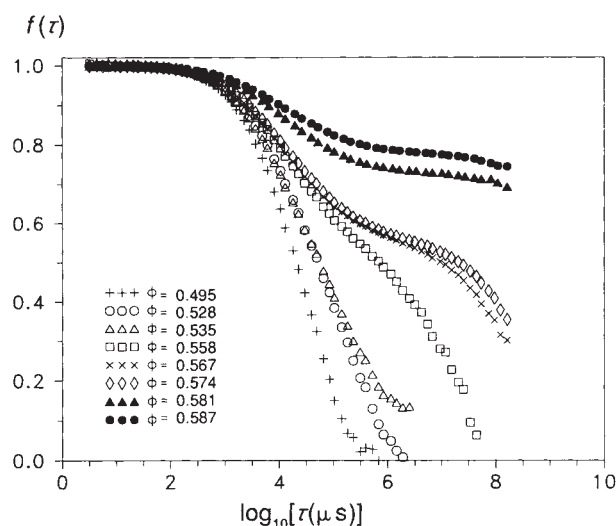


FIG. 2 Autocorrelation functions, $f(\tau)$, of particle concentration fluctuations versus the logarithm of the delay time τ , measured at an angle (80°) near the position of the main peak of the static structure factor. Except for the equilibrium fluid phase at freezing ($\phi_f=0.494$), measurements were made on the metastable fluid phases shortly after tumbling the samples but well before any crystals were evident.

ordered regions of particles. Below ϕ_g these highly asymmetric and thermodynamically unstable (due to their large surface to volume ratios) shear-induced nuclei dissipate by large-scale particle diffusion. Above ϕ_g , however, where dynamic light scattering indicates that large-scale diffusion is precluded, these structures remain frozen in the arrested metastable fluid. Particles adjacent to these shear-induced structures arrange themselves into registration through small-scale collective motions. This, in turn, provides additional free volume to the neighbouring particles aiding further growth of the plate-shaped crystals (Fig. 1) in the two most concentrated suspensions. This conjecture is given further support by the observation that the shape, orientation and growth rates of the crystals that develop in the kinetically arrested glass depend on the shear history; the vertically aligned needle-shaped crystals seen in the last two photographs in Fig. 1 developed within minutes of commencing a regular small-amplitude rocking motion.

As random close packing is approached, and even the smallest scale motions are gradually arrested⁶, undisturbed suspensions remain amorphous apart from some epitaxial crystal growth from the meniscus (see Fig. 1 of ref. 1). This growth is presumably aided by slow gravitational settling of the particles and registered on the shear-induced structures formed as the colloidal fluid drains down the walls of the cuvette after tumbling. One could further speculate that if formation of these shear-induced structures was prevented, by employing an alternative means of 'melting' or quenching at concentrations above ϕ_g , then the amorphous structure would persist indefinitely. □

Received 16 November 1992; accepted 1 March 1993.

1. Pusey, P. N. & van Meegen, W. *Nature* **320**, 340–342 (1986).
2. Pusey, P. N. in *Liquids, Freezing and the Glass Transition* (eds Hansen, J. P., Levesque, D. & Zinn-Justin, J. (North-Holland, Amsterdam, 1991).
3. Aastuen, D. J. W., Clark, N. A., Cotter, L. K. & Ackerson, B. J. *Phys. Rev. Lett.* **57**, 1733–1736 (1986).
4. Schätzel, K. & Ackerson, B. J. *Phys. Rev. Lett.* **68**, 337–340 (1992).
5. van Meegen, W. & Pusey, P. N. *Phys. Rev. A* **43**, 5429–5441 (1991).
6. van Meegen, W. & Underwood, S. M. *Phys. Rev. E* **47**, 248–261 (1993).
7. Paulin, S. E. & Ackerson, B. J. *Phys. Rev. Lett.* **64**, 2663–2666 (1980).
8. Hoover, W. G. & Ree, F. H. *J. chem. Phys.* **49**, 3609–3617 (1968).
9. Ackerson, B. J. *J. Rheol.* **34**, 553–590 (1990).

ACKNOWLEDGEMENTS. We thank B. Ackerson, A. Moodie and P. Pusey for discussions and P. Francis for technical assistance. Financial support for this work was provided by the Australian Research Council.

Scaling of the critical slip distance for seismic faulting with shear strain in fault zones

Chris Marone* & Brian Kilgore†

* Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, USA

† United States Geological Survey, Menlo Park, California 94025, USA

THEORETICAL and experimentally based laws for seismic faulting contain a critical slip distance^{1–5}, D_c , which is the slip over which strength breaks down during earthquake nucleation. On an earthquake-generating fault, this distance plays a key role in determining the rupture nucleation dimension⁶, the amount of premonitory and post-seismic slip^{7–10}, and the maximum seismic ground acceleration^{1,11}. In laboratory friction experiments, D_c has been related to the size of surface contact junctions^{2,5,12}; thus, the discrepancy between laboratory measurements of D_c ($\sim 10^{-5}$ m) and values obtained from modelling earthquakes ($\sim 10^{-2}$ m) has been attributed to differences in roughness between laboratory surfaces and natural faults⁵. This interpretation predicts a dependence of D_c on the particle size of fault gouge (breccia and wear material) but not on shear strain. Here we present experimental results showing that D_c scales with shear strain in simulated fault gouge. Our data suggest a new physical interpretation for the critical slip distance, in which D_c is controlled by the thickness of the zone of localized shear strain. As gouge zones of mature faults are commonly 10^2 – 10^3 m thick^{13–17}, whereas laboratory gouge layers are 1–10 mm thick, our data offer an alternative interpretation of the discrepancy between laboratory and field-based estimates of D_c .

We carried out experiments in a double-direct-shear testing apparatus (see ref. 18 for a description) in which two faults slide simultaneously (Fig. 1). We used quartz sand to simulate granular fault gouge and varied the initial layer thickness and particle size to assess their effect on D_c . Normal stress on the sliding surfaces was held constant at 25 MPa using a high-speed servocontroller and normal displacement was measured continuously to yield a record of gouge layer thickness during slip. To simulate the extreme roughness of natural fault surfaces, we sheared gouge between steel forcing blocks with grooves perpendicular to the sliding direction. This forces shear to occur within the layers and not at the layer boundaries¹⁹. A few trial tests and existing data¹⁹ indicate that the frictional behaviour and constitutive parameters determined from these experiments are the same as for experiments in which fault gouge is sheared between rocks of equivalent roughness.

The general friction characteristics and layer thickness changes observed were similar to those reported by previous workers^{2,20} (Fig. 1). Rapid compaction and strain hardening are followed by a transition to roughly linear layer thinning and shearing with constant coefficient of friction. The transition distance (~ 4 mm in Fig. 1) varies directly with layer thickness and initial gouge particle size. To investigate the constitutive properties of the gouge zone we imposed step changes in the remote slip velocity (between 1 and $10 \mu\text{m s}^{-1}$), producing changes in friction and in the rate of compaction with displacement (Fig. 1). For an increase in slip rate the coefficient of friction increases and the layer dilates, and for a decrease in slip rate, friction decreases and the layer compacts. Both friction and the rate at which layer thickness changes with displacement evolve following a change in slip rate (Fig. 1).

The parameter D_c is the characteristic distance of the exponential decay following the immediate friction change (Fig. 1). Because of the finite stiffness of the experimental apparatus, D_c must be obtained by modelling. We used a one-degree-of-

freedom spring-slider model to describe elastic interaction between the gouge layer and testing machine. This relation was coupled to a rate and single-state-variable constitutive law^{2,3}, and D_c was obtained from an iterative, least-squares inversion scheme²¹. A single-state-variable law typically simulates the data well, matching both the magnitude of the immediate friction change and the subsequent evolution. Details of the modelling and the other constitutive parameters will be reported elsewhere (C.M., manuscript in preparation). D_c values were obtained for 12 experiments using three layer thicknesses (0.7 mm, 2.1 mm and 4.2 mm) and three initial particle size distributions (fine, fractal and coarse; see Fig. 2 for details). Because of the large number of data points, only D_c values for velocity decreases are plotted (Fig. 2); velocity increases show the same trends.

The data for D_c against layer thickness (inset to Fig. 2) show two main features. First, in agreement with previous work^{2,20}, D_c scales directly with initial particle size, and for a given initial layer thickness and particle size, D_c decreases systematically with displacement and layer thinning. The rate of D_c reduction with thinning varies inversely with initial layer thickness and is roughly independent of initial particle size (inset to Fig. 2). Second, D_c is not in general a unique function of layer thickness. However, for a given displacement (lines in Fig. 2 inset) D_c scales roughly linearly with thickness at a rate of ~ 6 to $10 \mu\text{m mm}^{-1}$. This trend is lower than the trend of D_c against thickness for a given experiment, which ranges from about 21 to $43 \mu\text{m mm}^{-1}$, depending on initial thickness.

Differences in the D_c -thickness trends can be reconciled if we account for differences in shear strain between experiments (Fig. 2). Shear strain is calculated as $\int du/h$, where u is shear

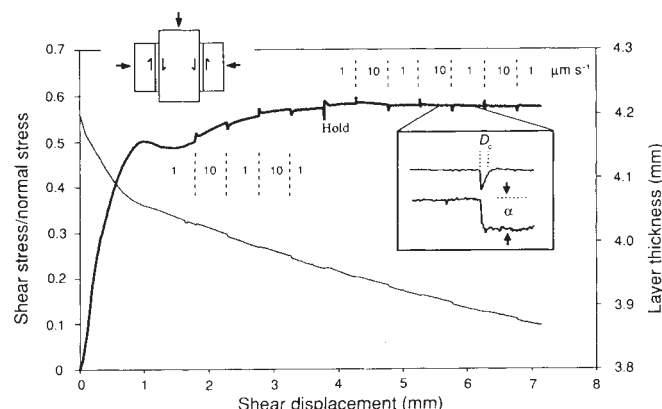


FIG. 1 Upper inset shows double-direct-shear configuration in which two gouge layers (stippled regions) are sheared simultaneously. Main plot shows normalized shear stress (bold curve) and layer thickness (of one layer) for fine quartz gouge. Beginning at ~ 2.0 mm displacement, the remote slip rate was cycled between 1.0 and $10.0 \mu\text{m s}^{-1}$. Inset shows enlargement of friction data (upper curve) and detrended layer thickness (a linear trend was removed from the thickness data). D_c is the evolution distance of friction following a change in remote slip rate. It represents a memory of friction for the past state of the slipping region, in this case a state produced by slip at a given velocity. Although D_c is obtained by modelling, it is possible to see directly from the data (main plot) that D_c decreases with displacement. The parameter α is the change in thickness due to the change in slip velocity. For granular fault gouge α is positive for an increase in velocity, representing dilation of the gouge layer¹⁹. Shear displacement has been corrected for elastic distortion of the experimental apparatus and layer thickness has been corrected for thinning which occurs as the gouge layer is offset²⁰. The gouge layer compacts during initial shear (up to ~ 1 mm) but subsequent thinning is due almost entirely to geometric thinning and localized strain along oblique Riedel shears³¹. 'Hold' denotes a point at which shear was momentarily (2 to 5 minutes) stopped while the control displacement transducer and data logger were reset. Data were collected every $0.25 \mu\text{m}$ of shear displacement.

displacement and h is layer thickness. The data show that D_c decreases systematically and nonlinearly with shear strain and that initially coarse particles yield the largest D_c values for all strains.

The reduction in D_c with shear strain could in principle be caused by comminution, given the particle-size dependence of D_c . But the average particle size for coarse gouge after shear is 50 – $100 \mu\text{m}$, whereas the ultra-fine particles start at 1 – $10 \mu\text{m}$. As D_c is larger for thick ultra-fine gouge than for thin coarse gouge (Fig. 2), the D_c reduction with strain cannot be a simple comminution effect. Furthermore, because coarse particles have a very narrow initial size range, and hence more open packing than the fractal and ultra-fine particles, one expects a higher comminution rate for the coarse particles and hence a larger change in D_c with strain. Instead, the change in D_c with shear strain is roughly independent of initial particle size distribution (Fig. 2). The change therefore cannot be explained by comminution.

The steeper D_c -thickness trend in a given experiment compared with that for a given displacement (Fig. 2 inset) implies that in an experiment D_c decreases because of layer thinning and a displacement effect such as comminution or strain localization. As argued above, if comminution were responsible we would expect the D_c -thickness relation for a given experiment to scale strongly with particle size, which it does not. On the other hand, if we assume that strain localizes with displacement, then the reduction in D_c during an experiment could reflect a progressive narrowing of the region over which strain is accommodated and processes controlling D_c operate. This interpretation is consistent with the form of the D_c -strain data, which, in agreement with theory and experimental evidence^{22,23}, would imply nonlinear thinning of the localized zone ('shear band').

Although we cannot directly measure shear band thickness, we assume that dilatancy and volumetric strain within the region accommodating shear provide an indirect measure of the volume of that region and, hence, shear band thickness²⁴. We define the parameter $\alpha = \Delta h / \Delta \log V$ (Fig. 1) as the change in layer thickness h on a change in slip velocity V (Fig. 3). α scales with particle size and shows a strong, positive correlation with D_c (inset to Fig. 3). Following the above discussion for D_c , the decrease of α with shear strain cannot be interpreted as a comminution effect, as it is roughly independent of particle size. Furthermore, comminution and subsequent tighter packing of the gouge particles should result in progressively larger dilation, and hence larger α , with shear for a given material volume, contrary to the observed trend (Fig. 3). On the other hand, under the assumption that the dilation (or compaction) magnitude scales with the volume of material participating in shear, the data of Fig. 3 imply progressive narrowing of the zone of strain accommodation. Together with the relation between D_c and α , this implies that the micromechanical processes controlling D_c operate within a zone of finite thickness, which narrows with increasing strain.

We offer the following physical interpretation for the relation between D_c and shear band thickness. For a shear band of thickness T and average grain diameter d , slip is distributed over $n = T/d$ particle-particle contacts. On a change in slip velocity at the shear zone boundary, each contact evolves with characteristic distance d_c and the shear band as a whole has $D_c = nd_c \chi$, where χ is a geometric factor to account for contact orientation. The parameter d_c is the characteristic friction distance for a single interface, which for bare rock surfaces is $\sim 1 \mu\text{m}$ (refs 2, 21). This can be related to grain diameter, d , as $d_c = d\zeta$, where ζ is a constant to account for the ratio of contact junction size to grain diameter and the displacement needed for sliding to develop fully at a contact²⁵. Combining these relations and the constants gives $D_c = nd\gamma = T\gamma$, which predicts a linear dependence of D_c on shear band thickness.

We may estimate the factor γ from our experiments if we

assume that under small strains shear is pervasive. The shear band then corresponds to the entire gouge thickness. As the minimum strain for localized shear is expected to increase with initial particle size^{20,26}, this assumption is most reasonable for small displacements and thick coarse gouge, for which $D_c = 31 \mu\text{m}$ and $T = 3.2 \text{ mm}$ (Fig. 2), yielding $\gamma \approx 0.01$. We can use our scaling relation to estimate D_c for natural faults by noting that recent work (ref. 15) indicates that most slip on mature faults occurs within the fault-zone core, in an ultracataclasite zone which corresponds to the zone of active shear. For an exhumed section of the southern San Andreas fault system, the thickness of the ultracataclasite zone is $T = 10 \text{ cm}$ (ref. 15), which from our scaling relation yields $D_c = 1 \text{ mm}$. This is in reasonable agreement with field and modelling estimates, which range from 0.1 mm (ref. 27) to $\sim 1 \text{ cm}$ (refs 1, 5–8, 29).

D_c has generally been interpreted in terms of the mechanics of contact junctions on surfaces free of gouge^{2,5}. Here, we have shown instead that D_c depends on the thickness of gouge that participates in shear. Our physical model, in which D_c for a shear band derives from d_c for the constituent particle-particle contacts, provides a mechanism for linking friction of fault gouge with friction of bare rock surfaces, and a means of generalizing the results of detailed friction studies of bare rock surfaces.

Our data imply that D_c for natural faults will depend on the nature of gouge and strain accumulation. For a fault zone which accumulates strain without significant wear and gouge zone thickening, the trend of D_c with shear strain should resemble our experimental data (Fig. 2), yielding a critical slip distance

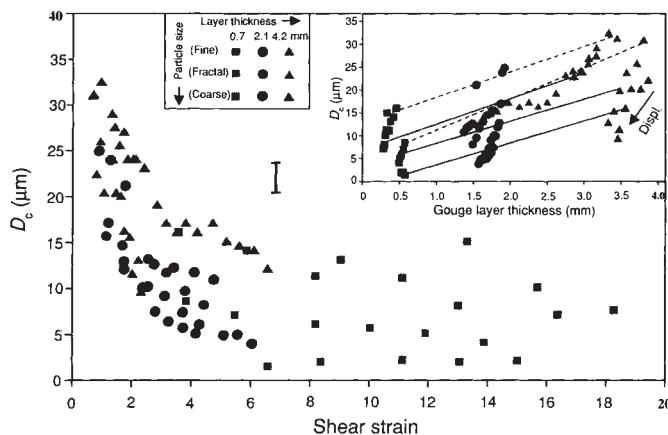


FIG. 2 D_c values from 12 experiments plotted against layer thickness (inset) and shear strain. Symbol colour and shape denote different initial particle sizes and layer thicknesses. Main plot: for a given initial particle size, D_c decreases systematically with shear strain. Inset: arrow shows increasing displacement during a given experiment. Several D_c values are obtained from each experiment, and layers thin with slip (for example, the red triangles are obtained from a single experiment, such as in Fig. 1). D_c is not a unique function of layer thickness, but lines connecting points of equal shear displacement indicate that for a given displacement and initial particle size D_c increases with thickness (dashed lines connect points for 2.3 mm displacement, solid lines connect 5.7 mm points). Error bar denotes resolution of D_c from modelling. Of the nine different initial configurations of particle size and layer thickness, we duplicated three sets (0.7-mm-thick coarse gouge; 2.1-mm ultra-fine gouge; and 4.2-mm coarse gouge) and checked several others with similar configurations (different normal stress, surface roughness or initial thickness). Experimental reproducibility was good. Starting gouge material: 'fine', commercially ground pure quartz powder, Silcosil 400 mesh, median and maximum diameter of 1.4 and $10 \mu\text{m}$, respectively³⁰; 'fractal', particle size distribution given by $N(n) = bn^{-D}$, where $N(n)$ is the number of particles of size n , b is a constant and D is the fractal dimension 2.6 (made, following ref. 20, by estimating the number of particles by weight and using size ranges from <45 to $710 \mu\text{m}$); 'coarse' Ottawa sand, ASTM C-190; particle diameter 600–800 μm .

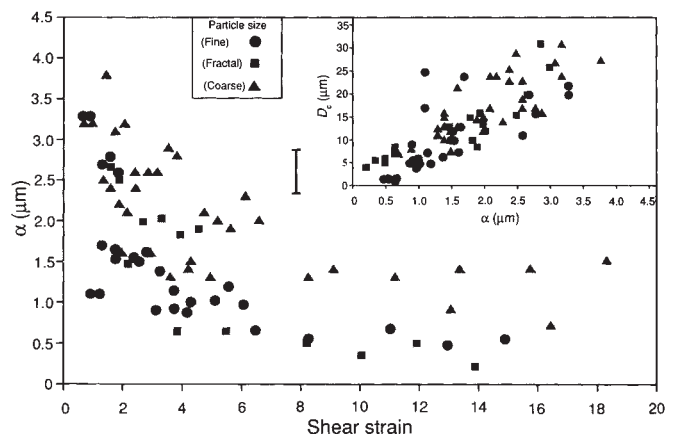


FIG. 3 Alpha is the change in layer thickness for a change in remote slip velocity (see inset to Fig. 1). Symbols denote different initial particle sizes. The data show a strong correlation with D_c (inset) and, within the resolution, a systematic variation with shear strain. Comparison with Fig. 2 indicates that α scales with shear strain and is independent of gouge thickness. Initially coarse particles yield the highest α for a given shear strain or displacement. We assume that α is a measure of the volume of gouge material participating in shear and hence the thickness over which strain is accommodated (shear band thickness). Thus, the relationship between α and D_c implies that D_c scales with shear band thickness. Error bar indicates measurement resolution.

independent of nominal gouge thickness. On the other hand, rapid gouge accumulation (via steady wear or inclusion of brecciated wall rock during earthquake rupture²⁸) or re-working of an existing gouge zone would produce a thick, low-strain fault zone and thus large D_c . This suggests that mature faults with thick gouge zones may have significantly higher D_c than faults with negligible net slip or those which otherwise have little accumulated gouge.

As D_c is an evolution distance, representing a memory of friction for past states of the slipping region, our data imply that faults with thick ultracataclasite zones will require larger slip to effect a given change in friction. On an earthquake fault, this implies greater precursory slip before rupture nucleation, and in general, a greater tendency for stable sliding, because instability is enhanced by rapid changes in frictional resistance. Thus mature faults with thick gouge zones and large D_c may show greater preseismic slip and have larger nucleation zones than faults with negligible gouge zones. □

Received 16 November 1992; accepted 9 March 1993.

1. Ida, Y. *J. geophys. Res.* **77**, 3796–3805 (1972); *Bull. seism. Soc. Am.* **63**, 959–968 (1973).
2. Dieterich, J. H. *J. geophys. Res.* **84**, 2161–2168 (1979); in *Mechanical Behaviour of Crustal Rocks*, *Geophys. Monogr. Ser. Vol. 24* (eds Carter, N. L., Friedman, M., Logan, J. M. & Sterns, D. W.) 103–120 (Am. Geophys. Un., Washington DC, 1981).
3. Ruina, J. *geophys. Res.* **88**, 10359–10370 (1983).
4. Andrews, D. J. *Bull. seism. Soc. Am.* **75**, 1–21 (1985).
5. Scholz, C. H. *Nature* **336**, 761–763 (1988).
6. Dieterich, J. H. in *Earthquake Source Mechanics* (eds Das, S., Boatwright, J. & Scholz, C.) 37–47 (Am. Geophys. Un., Washington DC, 1986); *Tectonophysics* **211**, 115–134 (1992).
7. Tse, S. T. & Rice, J. R. *J. geophys. Res.* **91**, 9452–9472 (1986).
8. Lorenzetti, E. & Tullis, T. E. *J. geophys. Res.* **94**, 12,343–12,361 (1989).
9. Marone, C., Scholz, C. H. & Bilham, R. *J. geophys. Res.* **96**, 8441–8452 (1991).
10. Stuart, W. D. *Pure appl. Geophys.* **126**, 619–641 (1988).
11. Aki, K. *J. geophys. Res.* **84**, 6140–6148 (1979); **92**, 1349–1355 (1987).
12. Rabinowicz, E. *J. appl. Phys.* **22**, 1373–1379 (1951); **27**, 131–135 (1956).
13. Feng, R. & McEvilly, T. V. *Bull. seism. Soc. Am.* **73**, 1701–1720 (1983).
14. Scholz, C. H. *Geology* **15**, 493–497 (1987).
15. Chester, F. M., Evans, J. P. & Biegel, R. L. *J. geophys. Res.* **98**, 771–786 (1993).
16. Michelini, A. & McEvilly, T. V. *Bull. seism. Soc. Am.* **81**, 524–552 (1991).
17. Montgomery, D. R. & Jones, D. L. *Geology* **20**, 55–58 (1992).
18. Linker, M. F. & Dieterich, J. H. *J. geophys. Res.* **97**, 4923–4940 (1992).
19. Marone, C., Raleigh, C. B. & Scholz, C. H. *J. geophys. Res.* **95**, 7007–7025 (1990).
20. Biegel, R. L., Sammis, C. G. & Dieterich, J. H. *J. struct. Geol.* **11**, 827–846 (1989).
21. Marone, C. & Cox, S. J. D. *Eos* **71**, 457 (1991).
22. Mühlhaus, H.-B. & Vardoulakis, I. *Géotechnique* **37**, 271–283 (1987).
23. Jaeger, H. M. & Nagel, S. R. *Science* **255**, 1523–1532 (1992).
24. Lambe, T. W. & Whitman, R. V. *Soil Mechanics* (Wiley, New York, 1969).

25. Boitnott, G. N., Biegel, R. L., Scholz, C. H., Yoshioka, N. & Wang, W. *J. geophys. Res.* **97**, 8965–8978 (1992).
 26. Marone, C. & Scholz, C. H. *J. struct. Geol.* **11**, 799–814 (1989).
 27. Power, W. L. & Tullis, T. E. *J. geophys. Res.* **97**, 15425–15435 (1992).
 28. Sibson, R. H. *Nature* **316**, 248–251 (1985); *Pure appl. Geophys.* **124**, 159–176 (1986).
 29. Rice, J. R. *J. geophys. Res.* (in the press).
 30. Wong, T.-F. & Zhao, Y. *Tectonophysics* **175**, 177–195 (1990).
 31. Scott, D. R., Marone, C. & Sammis, C. *J. geophys. Res.* (submitted).

ACKNOWLEDGEMENTS. The experiments were done at the U.S. Geological Survey, Menlo Park. C.M. thanks J. Dieterich for making his laboratory available for the experiments and for encouragement. We thank R. Jeanloz, C. Scholz and C. Sammis for helpful comments and suggestions on the manuscript. This work was supported by the NSF.

Variations in mercury deposition to Antarctica over the past 34,000 years

Grace M. Vandal*, William F. Fitzgerald*,
 Claude F. Boutron†‡ & Jean-Pierre Candelone†

* Department of Marine Sciences, University of Connecticut, Avery Point, Groton, Connecticut 06340, USA

† Laboratoire de Glaciologie et Géophysique de l'Environnement du CNRS, 54 rue Molière, Domaine Universitaire, BP 96, 38402 St Martin d'Hères cedex, France

‡ UFR de Mécanique, Université Joseph Fourier de Grenoble, Domaine Universitaire, BP 68 38041 Grenoble, France

POLAR ice contains a valuable record of past atmospheric mercury deposition, which can provide information about both the natural biogeochemical cycling of this toxic trace metal and the impact of recent anthropogenic emissions. But existing studies of mercury in polar ice and snow cores^{1–5} suffer from sample contamination and inadequate analytical procedures. Here we report measurements of mercury concentrations spanning the past 34,000 years from the Dome C ice core, Antarctica, using the stringent trace-metal clean protocols developed by Patterson and co-workers⁶. Although this record does not extend into the industrial period, it provides an important baseline for future attempts to identify anthropogenic mercury in Antarctic ice and snow. We find that mercury concentrations were strikingly elevated during the last glacial maximum (18,000 years ago), when oceanic productivity may have been higher than it is today⁷. As oceanic mercury emission is correlated with productivity^{8,9}, we suggest that this was the principal pre-industrial source of mercury to Antarctica; mercury concentrations in Antarctic ice might therefore serve as a palaeoproductivity indicator for the more distant past.

The historical record of mercury deposition contained in polar ice cores yields information on the natural biogeochemical mercury cycle and provides a way to assess the impact of modern anthropogenic emissions. The atmosphere is one main pathway for the distribution of mercury at the Earth's surface. For example, atmospheric mercury deposition greatly exceeds fluvial inputs to the oceans and is the principal input to many aquatic and terrestrial systems^{10–12}. Gaseous elemental Hg (Hg^0) is the dominant atmospheric species and has a residence time of about a year. Elemental Hg is oxidized in the atmosphere and removed by wet or dry deposition. Atmospheric deposition contains principally ionic and particulate Hg species^{11,12}. Modern human-related Hg emissions (coal combustion, waste incineration, smelting) to the atmosphere are comparable in magnitude to natural fluxes (oceanic emission, terrestrial biogenic volatiles, volcanoes)^{11,13,14}. The effects on the natural Hg cycle are of concern because Hg and many of its compounds are toxic and can be converted to more poisonous forms (such as monomethyl mercury) in aquatic systems^{15,16}. The surface waters of the equatorial Pacific Ocean are highly supersaturated with Hg^0 relative to atmospheric concentrations. Oceanic gas emission of

Hg^0 is a principal source of Hg to the atmosphere and the magnitude of the efflux is related to productivity^{8,9}.

We have measured Hg in 14 sections of the Dome C core. Dome C is located in central Antarctica (77° 39' S, 124° 10' E) at an elevation of 3,240 m. The core is 905 m long and covers the past 40,000 years¹⁷. Details of the trace metal clean procedures and mechanical decontamination techniques used in the treatment of the Dome C core can be found in ref. 6. Briefly, the core sections were processed at the California Institute of Technology in the following manner. Each section was mechanically cleaned by chiselling four to six successive veneers of ice from the exterior towards the interior of the core section under ultra-clean conditions. The veneer layers were collected and analysed separately to determine the trace metal content of each section as a function of radius. The samples were acidified with ultra-pure HNO_3 to make a 0.1% HNO_3 solution. Aliquots for the determination of lead were removed and the remaining sample kept frozen except to thaw for the removal of a sample aliquot for analysis. Mercury was determined in aliquots of the 14 sections used for the determination of lead and other constituents. In addition to the ice-core samples, numerous blanks were included to assess Hg contributions from processing the core sections, acidification and storage in polyethylene bottles. Samples were shipped frozen to the University of Connecticut in April 1991 for Hg analysis.

Mercury concentrations in the sections of the Dome C core are listed in Table 1. Aluminium, sodium and non-sea-salt (n.s.s.) SO_4 are also reported in Table 1. Non-sea-salt SO_4 is the SO_4 concentration corrected for the sea-salt component and represents the product of the atmospheric oxidation of reduced sulphur species. The core covers four main climatic stages, as inferred from the oxygen isotope profile¹⁷, and these are listed in chronological order: (1) the Holocene (0 to ~408 m depth), (2) the transition from the Last Ice Age to the Holocene (~408–548 m), (3) the Last Glacial Maximum (LGM; ~548–708 m)

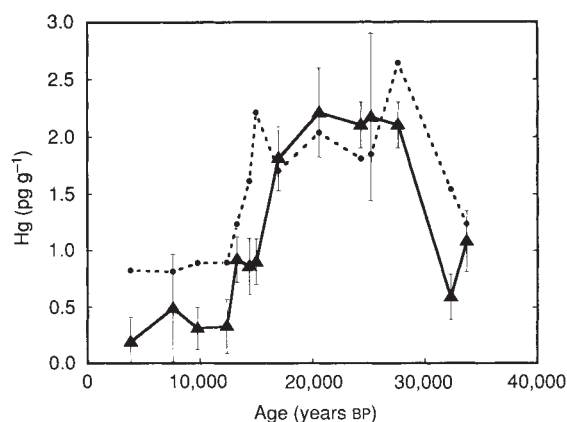


FIG. 1 a (▲) Mercury distribution as a function of age in the Dome C ice core. The Hg concentration was less than 1.0 pg g^{-1} during the Last Ice Age, then increased to 2.1 pg g^{-1} during the LGM, remaining elevated for ~10,000 years before decreasing during the glacial to interglacial transition to ~ 0.4 pg g^{-1} in the Holocene. Mercury was determined in the ice and snow samples by SnCl_2 reduction and collection on gold-coated sand with gas-phase detection by cold vapour atomic fluorescence spectroscopy^{22,29}. Analytical precision was generally $\pm 5\%$. A blank of $0.16 \text{ pg Hg g}^{-1}$ ice from the polyethylene sample bottles was detected and the measured Hg values corrected. Samples ranged from 3 to 40 g. The uncertainty in the reported concentrations associated with variations in analytical and storage blanks was $\pm 0.2 \text{ pg g}^{-1}$ for a 20-g sample. b, (---) The predicted distribution of Hg in the Dome C ice core based on the marine-derived n.s.s. SO_4 concentrations and the estimated Hg/S ratio for gaseous oceanic evasion. The predicted and measured values are remarkably close, except in the Holocene where the estimate is greater than the measured values. The Hg/S ratio used is representative of a productive ocean region and the ratio of their emissions during the Holocene may be lower.

TABLE 1 Mercury and ancillary measurements

Depth (m)	Age (yr BP)	Inner core Hg concentration (pg g^{-1})	Plateau for Hg [†]	Na [‡] (ng g^{-1})	Al [‡] (ng g^{-1})	n.s.s. SO ₄ [‡] (ng g^{-1})
172.8	3,850	0.19	—	13	0.95	68
300.6	7,600	0.49	+	16	1.6	67
373.9	9,800	0.31	+	7.5	1.4	73
451.9	12,400	0.33	+	35	2.4	73
476.3	13,300	0.92	0	22	1.5	101
500.5	14,400	0.86	+	52	7.2	133
527.2	15,000	0.90	0	83	35	182
545.1	17,000	1.81	+	65	28	140
602.2	20,600	2.21	—	101	32	167
658.2	24,300	2.14	0	84	48	149
670.5	25,200	2.17	+	80	50	152
704.2	27,600	2.14	0	102	77	217
775.7	32,300	0.59	+	55	6.0	127
796.9	33,700	1.08	—	59	5.0	102

* Reported values have been blank-corrected.

† Potential sample contamination associated with the coring process has been assessed using outside to inside Hg concentration profiles obtained from analysis of veneer layers for a section. Low and consistent Hg levels in the inner layers relative to the outer veneer which contacts the corer is an indication that contamination did not penetrate the inner core of a section. (+) The variation of Hg concentration with radius indicated that the inner core sample was representative of the natural level. The outside veneer layers had significantly higher concentrations than the inner veneer layers for three (300, 500, 775 m) of these seven profiles, whereas the Hg content of the veneer layers was uniform and low for the other four sections (374, 451, 545, 670 m). (—) Profiles showed more variability and the inner core value is taken as an upper limit of the natural Hg concentration. (0) Only the inner core was analysed.

‡ Na, Al and SO₄ determined previously³⁰.

and (4) the early Last Ice Age stage (~708 to bottom of the core). Figure 1 shows the Hg profile for the Dome C ice core. The Hg concentration in the ice was less than 1.0 pg g^{-1} at 34,000 years before present (BP), then increased to 2.1 pg g^{-1} and remained elevated between 28,000 and 18,000 BP, coinciding with the LGM. The Hg concentration declined between 17,000 and 13,000 BP, during the Last Ice Age/Holocene transition, and levelled off to less than 0.5 pg g^{-1} during the Holocene.

Volcanic, continental and sea-salt contributions to Hg in the ice core are small. Estimates of the mean global Hg/S mass ratio in volcanic gases are between 1×10^{-6} (ref. 18) and 10×10^{-6} (ref. 19). The computed volcanic contribution ranges from 0.03 pg g^{-1} during the Holocene to 0.09 pg g^{-1} during the LGM, assuming that 13% of the n.s.s. SO₄ in the ice is of volcanic origin⁶ and using the upper estimate for the Hg/S mass ratio in volcanic emissions. Soil dust, shown to be the main source of other trace metals (Pb, Cu, Cd and Zn) to the Dome C ice core^{6,20}, is not an important contributor of Hg to the Antarctic. The average Hg/Al ratio for crustal material is 1×10^{-6} (ref. 21) and we estimate a contribution of Hg from soil dust of 0.001 to 0.077 pg g^{-1} on the basis of the measured Al concentrations for the sections. The sea-salt contribution is insignificant and is estimated at less than $1 \times 10^{-5} \text{ pg Hg g}^{-1}$ of ice on the basis of the sodium concentrations and a Hg/Na ratio for open ocean water of 1×10^{-12} (ref. 22). The combined contribution of soil dust, volcanoes and sea-salt ranges from 0.03 pg g^{-1} during the Holocene to 0.17 pg g^{-1} during the LGM, accounting for 5–10% of the Hg measured in the ice. Terrestrial biogenic emissions currently contribute ~20% of the natural Hg flux to the atmosphere¹³. We argue, however, that this source was not a significant contributor to Antarctic Hg deposition, as Antarctic ice-core studies of trace gases (CH₄ and N₂O) with important terrestrial biogenic sources show lower concentrations during the LGM²³, whereas Hg increased markedly during this period.

The emission of gaseous mercury from productive ocean regions has been postulated as an important source of Hg to the atmosphere⁵. The magnitude of the elemental Hg emission from the Southern Ocean is not currently known. However, there is information available about dimethyl sulphide (DMS), n.s.s. SO₄, and the atmospheric sulphur cycle for this region. There are several important sources of n.s.s. SO₄ to the Antarctic,

of which oceanic emission and subsequent oxidation of DMS is considered to be primary, accounting for ~80% of the total n.s.s. SO₄ in aerosols²⁴. Although the processes that lead to the emission of DMS and Hg from the oceans are probably different, both are related to ocean productivity^{8,25}. Estimates of the ocean contribution of Hg to Dome C were made on the basis of the marine derived n.s.s. SO₄ in the ice and using a Hg/S mass ratio of 4.4×10^{-5} , taken from their respective effluxes from the equatorial Pacific Ocean^{8,25}. These results are presented in Fig. 1. The predicted values are close to the measured concentrations. We suggest that oceanic emission of gaseous Hg was the principal source of Hg to the Antarctic during pre-industrial times. Further evidence is provided by the covariation of Hg and methanesulphonate (MSA), a by-product of the atmospheric oxidation of DMS. Whereas the concentration of n.s.s. SO₄ (149 – 217 ng g^{-1}) was a factor of 2–3 higher during the LGM than during the Holocene (67 – 73 ng g^{-1}), the MSA concentrations were 5–6 times higher in ice from the LGM (~ 15 ng g^{-1}) than in ice from the Holocene ice (~ 3 ng g^{-1})²⁶. The variations of MSA in the Dome C core correspond with those of Hg which was also five times greater in LGM ice (~ 2.0 pg g^{-1}) than in Holocene ice (~ 0.4 pg g^{-1}). The higher concentrations during the LGM may be attributed to enhanced oceanic production, more efficient gas exchange and/or increased atmospheric transport to the Antarctic.

The depositional flux of Hg over time was calculated from the water accumulation rate¹⁷. At Dome C the Hg depositional flux was ~ $1.3 \text{ pg cm}^{-2} \text{ yr}^{-1}$ during the Last Ice Age, increased to $3.1 \text{ pg cm}^{-2} \text{ yr}^{-1}$ during the LGM and decreased to $0.9 \text{ pg cm}^{-2} \text{ yr}^{-1}$ during the Holocene. The threefold increase in Hg deposition to Dome C during the glacial period is consistent with the estimates of enhanced marine biogenic productivity in upwelling and sub-polar areas during this time⁷. Recent work²⁷ suggests that productivity in the nutrient-rich waters of the Southern Ocean is limited by the availability of Fe. Elevated oceanic Hg emissions during the LGM would be consistent with increased productivity in the Southern Ocean due to enhanced atmospheric transport of crustally derived Fe. In addition, enhanced upwelling⁷ and continental erosion²⁸ during the LGM would magnify the Hg efflux from the surface waters by supplying additional Hg for reduction and emission. The higher concentration of aerosols in ice from the LGM has been attributed

to increased poleward atmospheric transport, stronger winds and a more arid climate²⁸. This climatic effect would intensify Hg deposition to the Antarctic during the LGM.

The relative importance of terrestrial and oceanic Hg sources to the Arctic environment is currently unknown. It is likely, however, that terrestrial sources will be more important there as the ratio of the land mass to ocean is higher in the Northern Hemisphere than in the Southern Hemisphere. Whether

oceanic Hg emission and subsequent terrestrial deposition might serve as an indicator of current and past productivity must be evaluated. During the LGM, marine Hg emissions and deposition at Dome C may have been enhanced by strong winds and poleward atmospheric transport. Modern air-sea exchange studies of Hg production and emission in the Southern Ocean coupled with atmospheric depositional studies on the Antarctic continent would be helpful. □

Received 29 October 1992; accepted March 1993.

- Weiss, H. V., Koide, M. & Goldberg, E. D. *Science* **174**, 692-694 (1971).
- Appelqvist, H., Jensen, K. O., Sevel, T. & Hammer, C. *Nature* **273**, 657-659 (1978).
- Carr, R. A. & Wilkniess, P. E. *Science* **181**, 843-844 (1973).
- Herron, M. M. *et al.* in *Isotopes and Impurities in Snow and Ice*, 98-102 (Int. Assoc. Sci. Hydrol., 1977).
- Murozumi, M., Nakamura, S. & Yoshida, Y. *Mem. natn. Inst. Polar Res.* **7**, 255-263 (1978).
- Bouton, C. F. & Patterson, C. C. *Nature* **323**, 222-225 (1986).
- Sarnthein, M., Winn, K. & Zahn, R. in *Abrupt Climatic Change* (eds Berger, W. H. & Labeyrie, L. D.) 311-337 (Reidel, Holland, 1987).
- Kim, J. P. & Fitzgerald, W. F. *Science* **231**, 1131-1133 (1986).
- Mason, R. P. & Fitzgerald, W. F. *Nature* **347**, 457-459 (1990).
- Fitzgerald, W. F. in *The Role of Air-Sea Exchange in Geochemical Cycling* (ed. Buat-Menard, P.) 363-408 (Reidel, Boston, 1986).
- Fitzgerald, W. F. in *Chemical Oceanography*, Vol. 9 (eds Riley, J. P. & Chester, R.) 151-186 (Academic, London, 1989).
- Lindqvist, O. *Water Air Soil Poll.* **55** (1981).
- Nriagu, J. O. *Nature* **338**, 47-49 (1989).
- Nriagu, J. O. & Pacyna, J. M. *Nature* **333**, 134-139 (1988).
- Fitzgerald, W. F. & Clarkson, T. W. *Envir. Health Pers.* **96**, 159-166 (1991).
- Beiger, J. & Jernelöv, A. in *The Biogeochemistry of Mercury in the Environment* (ed. Nriagu, J.) 203-210 (Elsevier-North Holland, Amsterdam, 1979).
- Jouzel, J. *et al.* *Quat. Res.* **3**, 135-150 (1989).
- Patterson, C. C. & Settle, D. M. *Geochim. Cosmochim. Acta* **51**, 675-681 (1987).
- Varekamp, J. C. & Buseck, P. R. *Appl. Geochem.* **1**, 65-73 (1986).
- Batfai, F., Bouton, C. F. & de Andelis, M. *Nature* **337**, 544-546 (1989).
- Taylor, S. R. *Geochim. cosmochim. Acta* **28**, 1273-1285 (1964).
- Gill, G. A. & Fitzgerald, W. F. *Mar. Chem.* **20**, 227-243 (1987).
- Leuenberger, M. & Siegenthaler, U. *Nature* **360**, 449-451 (1992).
- Prospero, J. M., Savoie, D. L., Saltzman, E. L. & Larsen, R. *Nature* **350**, 221-223 (1991).
- Cline, J. D. & Bates, T. S. *Geophys. Res. Lett.* **10**, 949-952 (1983).
- Saigne, C. & Legrand, M. *Nature* **330**, 240-242 (1987).
- Martin, J. H., Gordon, R. M. & Fitzwater, S. E. *Nature* **345**, 156-158 (1990).
- Petit, J. R., Briat, M. & Royer, A. *Nature* **293**, 391-393 (1981).
- Bloom, N. & Fitzgerald, W. F. *Analyt. chim. Acta* **208**, 151-161 (1988).
- Bouton, C. F., Patterson, C. C. & Bankov, N. I. *Earth Planet. Sci. Lett.* **102**, 248-259 (1990).

ACKNOWLEDGEMENTS. This work was supported in France by the Ministère de l'Environnement, the University of Grenoble, the Terres Australes et Antarctiques Françaises and the Expéditions Polaires Françaises; and in the USA by the Division of Polar Programs of the U.S. National Science Foundation. We thank C. C. Patterson for development of the ultra-clean procedures used here and for his participation in decontamination of the samples.

Flightless bird from the Cretaceous of Mongolia

Perle Altangerel*, Mark A. Norell†‡, Luis M. Chiappe† & James M. Clark†

* Mongolian Museum of Natural History, Ulan Bator 46, Zaluchidiin Gudamz-1, Mongolia

† Department of Vertebrate Paleontology, American Museum of Natural History, 79th Street @ CPW, New York, New York 10024, USA

THE Late Cretaceous rocks of Mongolia have produced unusual and phylogenetically important dinosaurs^{1,2}. Here we report a startling new example, *Mononychus olecranus* gen. et sp. nov., an avialian theropod dinosaur with a short, robust forelimb possessing a single stout claw. Several features, including a carinate sternum and reduced fibula, suggest that *Mononychus olecranus* is more closely related to modern birds than is *Archaeopteryx lithographica*. The two skeletons are among the best preserved fossils known of a primitive bird, and emphasize the complexity of the morphological transformation from nonavialian theropods to modern birds. The occurrence of such a primitive bird in the Late Cretaceous reflects the paucity of Mesozoic bird fossils and suggests that the early radiation of avialians is only beginning to be sampled.

Two articulated specimens were recently discovered; the holotype is from slightly younger beds and is about twice as large as the older specimen. Each comprises unquestionably associated parts of a single animal.

Theropoda
Maniraptora³
Avialae³

Metornithes, taxon nov.

Discussion. *Archaeopteryx* is traditionally included within Aves^{4,5} but a recent proposal³ restricts Aves to the crown group of extant birds, thus excluding *Archaeopteryx*. The group includ-

ing *Archaeopteryx* and all other birds is termed Avialae³. *Mononychus olecranus* belongs to a group within Avialae not including *Archaeopteryx*, the Metornithes (Fig. 1). We use the term 'bird' for avialians.

Etymology. *Meta* (Greek), among; *ornithes* (Greek), bird.

Diagnosis. See Fig. 1.

Mononychus olecranus, gen. et sp. nov.

Holotype. Mongolian Geological Institute (MGI) 107/6; collected by 1987 Soviet-Mongolian Paleontological Expedition. Includes posterior part of skull, most precaudal vertebrae, all four limbs, thoracic girdle and portions of ilium and pubis.

Etymology. *Mono* (Latin), one, *onyx* (Greek), claw; and *olekranon* (Greek), elbow head.

Locality and horizon of holotype. Upper Cretaceous (Maestrichtian^{6,7}) Nemegt Formation, Bugin Tsav, Mongolia.

Referred specimen. MGI N100/99, a skull fragment and articulated postcranial skeleton lacking cervical vertebrae, thoracic girdle and proximal forelimb elements. Collected by 1992 Mongolian-American Museum of Natural History Expedition from the Upper Cretaceous (Campanian^{6,7}) Djadokhta Formation, Tugrueen Shireh, Mongolia. All elements in common with the holotype are nearly identical.

Diagnosis. See Fig. 1.

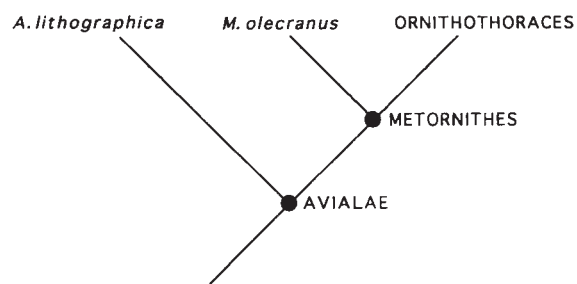
Description. Between the two specimens, nearly every skeletal element is represented (Fig. 2). A furcula was not found with either specimen.

The well preserved braincase displays caudal and rostral tympanic recesses. The single quadrate articular facet lies anterior to the otic recess. The right maxilla preserves the anterior edge of the antorbital fossa, which, unlike that in *Archaeopteryx* and nonavialian theropods, lacks an accessory fenestra. The ventral margin of the maxilla lacks alveoli.

A tiny tooth was recovered inside the skull, presumably from the premaxilla or mandible (Fig. 3). The tooth is sharply pointed, laterally compressed, with anterior and posterior carinae that broaden from the apex toward the constricted base. Constricted tooth bases are also present in *Archaeopteryx*⁸ and *Archaeornithoides*⁹, but they are conical and lack carinae. Carinate teeth are present in some avialians (such as *Ichthyornis*).

‡ To whom correspondence should be addressed.

FIG. 1 Cladogram of proposed relationships of *Mononychus olecranus*, gen. et sp. nov., and characters diagnosing groups. Metornithes is diagnosed by the following unambiguous characters¹⁰: ossified large and longitudinally oriented rectangular sternum; ossified sternal carina; prominent antitrochanter on ilium; undivided femoral trochanteric crest; fibula does not reach tarsus. The following diagnostic characters are ambiguous owing to lack of information or conflicting characters in closely related taxa¹⁰: carpometacarpus; pubic apices not in contact; wide vertebral foramen; femoral distal condyles nearly confluent below popliteal fossa; pubis without foot (present in *Sinornis*). *Mononychus olecranus* is diagnosed by the following unambiguous characters: edentulous maxilla (convergent with Aves); pronounced pectoral crest of humerus; single distal condyle of humerus; ventral tubercle of humerus pronounced; extremely short shafts of ulna and radius; very long olecranon process of ulna; carpometacarpus massive, short, quadrangular with no intermetacarpal space; manus digit I much larger than digits II and III; claw of manus digit I robust; coracoid not expanded ventrally (reversal to primitive maniraptoran condition); sternal carina thick; one posterior dorsal vertebra biconvex and xiphisternum procoelous (convergent with *Patagopteryx*)^{10,11}; zygapophyses, costal fossa, and transverse process of anterior dorsal vertebrae on same level; sharply keeled posterior xiphisternum centra; elongate haemal arches; ischium extremely slender; medial cnemial crest on tibiotarsus (convergent with Ornithurae)¹⁰; metatarsal III limited to distal third, triangular in cross-section (convergent with several nonavianian



theropods)¹⁸. Absence of a pubic foot is an ambiguous character. Ornithothoraces¹⁰, including all extant birds, is diagnosed by the following unambiguous characters: strut-like coracoid; scapulocoracoid articulation well below shoulder end of coracoid; coracoid forming sharp angle with scapula at level of glenoid cavity; scapula tapers caudally; humerus shorter than ulna; ulnar shaft much thicker than radial shaft. The following diagnostic characters are ambiguous: ventral ribs (gastralia) absent; U-shaped furcula (interclavicular angle smaller than 90 degrees)²³.

Cervical vertebrae are longer than in *Archaeopteryx*. The vertebral foramen is large compared with nonavianian theropods. Cervical ribs are not fused to the centra; well developed ventral processes are present on at least one anterior dorsal. As in *Patagopteryx*^{10,11}, one posterior dorsal is biconvex; more cranial vertebrae are opisthocoeleous. The xiphisternum (incorporating seven vertebrae) and caudal vertebrae are procoelous. Posterior xiphisternum centra are compressed transversely and keeled. Nineteen caudal vertebrae are preserved in N100/99, but some distal elements are missing. Long haemal arches articulate with the proximal caudals. Caudal zygapophyses are long but do not span adjacent vertebrae. Dorsal vertebrae lack the hypophenium and xiphisternum typical of nonavianian theropods. No vertebrae are heterocoelous.

The short humerus has a pronounced pectoral crest and a single distal condyle (Fig. 4). The condyle is on the cranial surface, as are both humeral condyles of extant birds¹⁰. The ventral humeral tubercle is pronounced. Radial and ulnar shafts are short and broadly abut proximally. The olecranon process is exceptionally long.

The short, massive carpometacarpus comprises three metacarpals and at least one distal carpal. Its proximal surface is remarkably similar to the 'semilunate' carpal of nonmetornithine

maniraptorans^{12,13}, indicating fusion of this element to the metacarpals. Unlike nonmetornithines, this 'semilunate' carpal is restricted to metacarpal I. As in *Archaeopteryx* and nonavianian theropods, metacarpal I is broadest, but metacarpals II and III are unusually small. Minute articulation facets on metacarpals II and III suggest tiny digits may have been present. The stout claw of the large first digit is unlike the delicate, recurved manus claws of most theropods.

The scapula and coracoid are similar to those in nonmaniraptoran theropods. The coracoid is not as deep as in dromaeosaurids, *Archaeopteryx* and, to a greater extent, modern birds. A bicapital tubercle is absent from the coracoid. The sternum is unlike the flat, quadrangular sternum of nonavianian Maniraptora¹⁴ in being carinate and longer than wide, but it is stouter than in other carinate birds.

The pelvic girdle and hindlimb are gracile. The ilium has a flat, horizontal postacetabular wing with a shallow brevis fossa. A distinct antitrochanter is present, a feature poorly developed or absent in *Archaeopteryx* and nonavianian theropods. As in extant birds the caudally oriented pubis and ischium lack symphyseal; unlike in all theropods, they are extremely slender and in contact throughout their length. The pubis lacks the distal foot typical of nonmetornithine theropods and *Sinornis*¹⁵.

FIG. 2 Composite skeletal reconstruction of *Mononychus olecranus*, gen. et sp. nov., in left lateral view, based on the holotype and N100/99. Unpreserved elements are filled in black. A, biconvex dorsal vertebra; B, opisthopubic pelvis; C, short, distally tapering fibula with large laterally projecting iliofibularis tubercle (arrow); D, proximal view of left tibial articular surface showing pair of cnemial crests (arrows); and E, rectangular, carinate sternum in anterior (top) and ventral (bottom) views. A, D and E from the holotype, B and C are from N100/99. Insets not to scale.

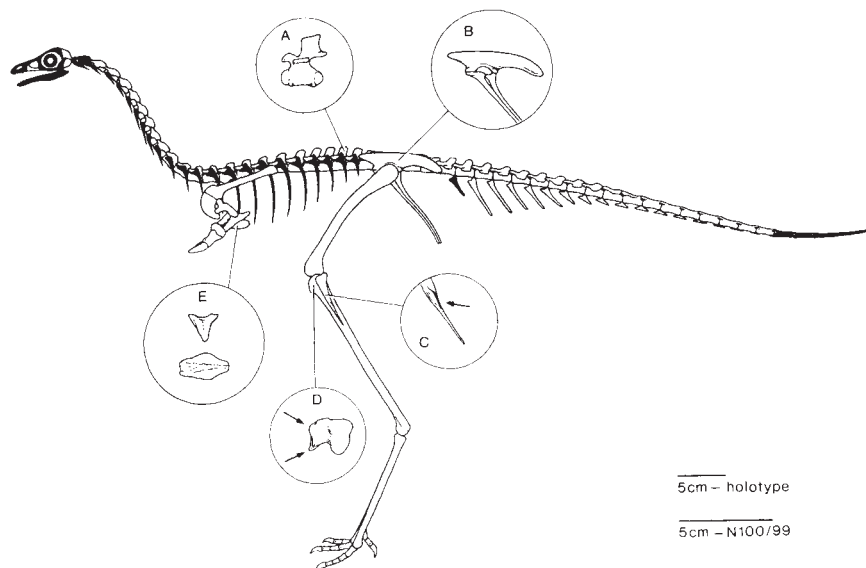
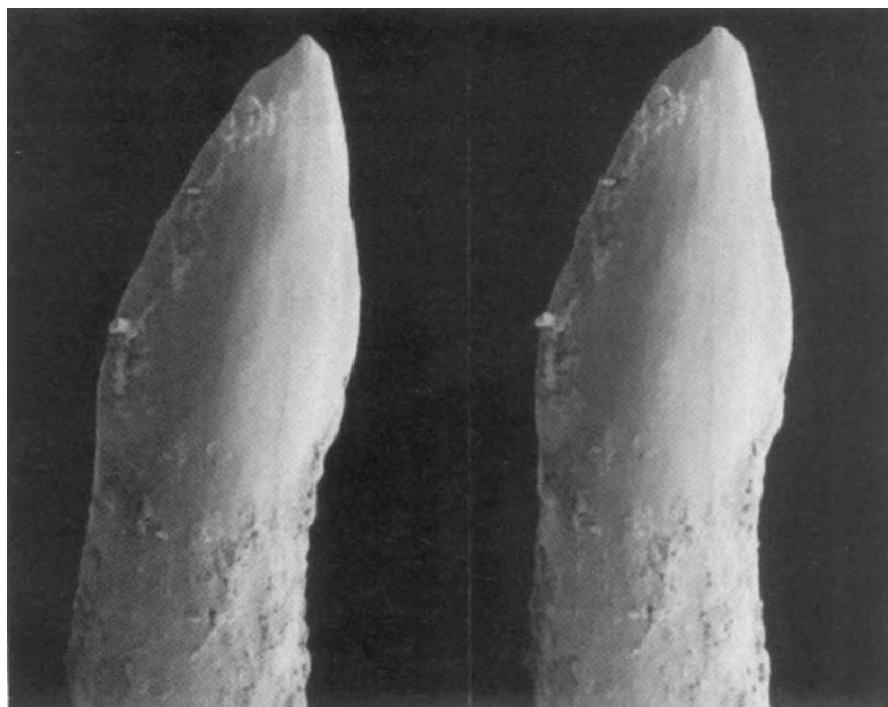


FIG. 3 Electron stereomicrograph of isolated tooth of *Mononychus olecranus*, gen. et sp. nov., holotype. Magnification $\times 91$.



Unlike *Archaeopteryx* and nonavianian theropods, but as in extant birds, the anterior and greater trochanters are fused, forming a trochanteric crest on the proximal end of the femur. The popliteal fossa is bounded distally by projections from both condyles, approaching the condition present in more advanced birds¹⁰.

The tibia and proximal tarsals of the holotype are partially fused into a tibiotarsus, but those of N100/99 are unfused. A weak medial cnemial crest is present, as in ornithurine birds¹⁰. A large astragalar ascending process is present. Like more advanced birds, the fibula does not contact the tarsus.

The metatarsals are unfused, unlike in *Archaeopteryx*¹⁶ and extant birds, but as in *Iberomesornis*¹⁷. Metatarsals II and IV are elongate, V is reduced to a tiny splint, and I is similar in size to that of nonavianian Maniraptora. Only the distal end of metatarsal III is developed; the bone tapers proximally less than half the length of II and IV, as in some nonavianian theropods¹⁸. The digital formula is 2-3-4-5-0 and, unlike dromaeosaurid theropods, the second digit is not retractable and lacks a sickle-shaped claw¹³.

Discussion. Five unambiguous derived characters indicate *Mononychus* is more closely related to extant birds than is *Archaeopteryx* (Fig. 1). All are unique to metornithines among theropods. Two characters contradict this hypothesis: attenuated metatarsal III and short, unexpanded coracoid. *Mononychus* lacks six unambiguous characters diagnosing more advanced birds (Ornithothoraces¹⁰).

The highly modified forelimb of *Mononychus* is similar to that of digging animals. The large pectoral process of the humerus, long olecranon process of the ulna, short, massive forelimb elements and carpometacarpus, and sturdy claw suggest extremely powerful functional capabilities during adduction. The short forelimb and long, gracile hindlimb are, paradoxically, incongruous with a burrowing habitus.

Despite its unique forelimb, *Mononychus* offers important evidence for bird evolution. In many other ways it is 'transitional' between primitive maniraptorans and modern birds. Furthermore, the presence of derived pelvic and hindlimb structures, combined with a primitively long tail, indicates tail shortening occurred after the acquisition of these features.

The forelimb of *Mononychus*, which is clearly incapable of flight, complicates interpretation of the evolution of avialian flight. Flight capabilities in *Archaeopteryx* are controversial, but most agree it possessed some capacity for active flight⁵. One interpretation is that flight is primitive for Avialae and was lost in the lineage leading to *Mononychus*. An equally parsimonious alternative is flight arose separately in *Archaeopteryx* and Ornithothoraces. It is impossible to choose between these two alternatives without invoking assumptions about the evolutionary process.

Many questions surround bird origins. The exemplars of early avialian evolution are the spectacular specimens of *Archaeopteryx*, but these are preserved in essentially two dimensions, and critical features are difficult to interpret⁵. Other specimens germane to this problem are incomplete⁹, crushed^{15,17} or disassociated¹⁹⁻²¹. This limits examination of many aspects of bird origins. For example, speculations about range of forelimb movement in primitive avialians^{15,22} are questionable when based on crushed specimens. The specimens of *Mononychus* are the first definitely associated, three-dimensional representatives of this critical phase of avialian evolution, and provide direct evidence for primitive character conditions near the base of the avian clade.

The discovery of *Mononychus* emphasizes the complexity of early bird evolution, highlighting our ignorance of the transition

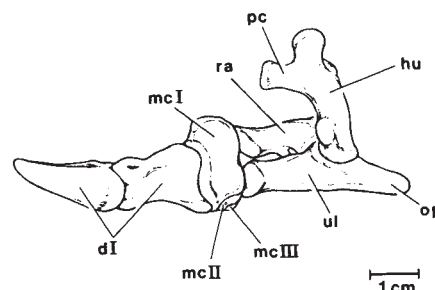


FIG. 4 Forelimb of *Mononychus olecranus*, gen. et sp. nov., in dorsal view. Abbreviations: dl, digit I; hu, humerus; mcI-III, metacarpals within carpometacarpus; op, olecranon process; pc, pectoral crest; ra, radius; ul, ulna.

from large, cursorial theropods to modern birds and documenting unexpected morphological diversity among primitive birds. □

Received 7 January; accepted 11 March 1993.

1. Andrews, R. C. *The New Conquest of Central Asia. Natural History of Central Asia Vol. I* (The American Museum of Natural History, New York, 1932).
2. Weishampel, D. B., Dodson, P. & Osmolska, H. *The Dinosauria* (University of California, Berkeley, 1990).
3. Gauthier, J. *Mem. Calif. Acad. Sci.* **8**, 1–46 (1986).
4. Carroll, R. *Vertebrate Paleontology and Evolution* (Freeman, New York, 1988).
5. Hecht, M. K., Ostrom, J. H., Viohl, G. & Wellnhofer, P. *The Beginnings of Birds* (Proc. Int. Archaeopteryx Conf., Freunde des Jura-Museums, Eichstätt, 1985).
6. Jerzykiewicz, T. & Russell, D. A. *Cret. Res.* **12**, 345–377 (1991).
7. Lillegraven, J. A. & McKenna, M. C. *Am. Mus. Novitates* **2840**, 1–68 (1986).
8. Martin, L. D. in *The Beginnings of Birds* (eds Hecht, M. K., Ostrom, J. H., Viohl, G. & Wellnhofer, P.) 177–183 (Proc. Int. Archaeopteryx Conf., Freunde des Jura-Museums, Eichstätt, 1985).
9. Elzanowski, A. & Wellnhofer, P. *Nature* **359**, 821–823 (1992).
10. Chiappe, L. M. in *3rd Symp. Soc. Avian Paleontology and Evolution* (ed. Peters, D. S.) (Senckenburgiana, in the press).
11. Chiappe, L. M. thesis, Univ. Nac. Buenos Aires (1992).
12. Ostrom, J. H. *Biol. J. Linn. Soc.* **8**, 91–182 (1976).
13. Ostrom, J. H. in *The Dinosauria* (eds Weishampel, D. B., Dodson, P. & Osmolska, H.) 269–279 (University of California, Berkeley, 1990).
14. Barsbold, R. *Sovm. Sov.-Mong. Paleontol. Eksped. Trudy* **19**, 1–117 (1983).
15. Sereno, P. & Rao, C. *Science* **255**, 845–848 (1992).
16. Wellnhofer, P. *Archaeopteryx* **6**, 1–30 (1988).
17. Sanz, J. L. & Bonaparte, J. F. in *Papers in Avian Paleontology Honoring Pierce Brodkorb* (ed. Campbell, K. E. Jr) 39–49 (Nat. Hist. Mus. Los Angeles, Sci. Ser. 36, Los Angeles, 1992).
18. Holtz, T. R. *J. Vert. Paleont.* **12** (suppl.), 33A (1992).
19. Chatterjee, S. *Phil. Trans. R. Soc. B332*, 277–346 (1991).
20. Kurzanov, S. M. *Sovm. Sov.-Mong. Paleontol. Eksped. Trudy* **31**, 5–95 (1987).
21. Ostrom, J. H. *Archaeopteryx* **5**, 113–114 (1987).
22. Gauthier, J. A. & Padian, K. in *The Beginnings of Birds* (eds Hecht, M. K., Ostrom, J. H., Viohl, G. & Wellnhofer, P.) 185–197 (Proc. Int. Archaeopteryx Conf., Freunde des Jura-Museums, Eichstätt, 1985).
23. Chiappe, L. M. *Alcheringa* **15**, 333–338 (1991).

ACKNOWLEDGEMENTS. We thank Barsbold R., Dashzeveg D., Namsarai B., M. Novacek, M. C. McKenna, M. Ellison, A. Davidson, L. Witmer, J. Cracraft, J. Ostrom, J. Davis and P. Sereno. Supported in part by a grant from IREX to M. Novacek and a grant from the P. M. McKenna Foundation to the American Museum.

TABLE 1 Telemetered *Nautilus* mantle cavity pressures and conversions

Situation	Pressure (Pa)	Speed (cm s ⁻¹), angle		O ₂ consumption (ml kg ⁻¹ h ⁻¹)*
		Estimated	Measured	
Maximum (30 s, release)	178	11.0		80
Climb (5 min, release)	70	4.0	4.2, 70°	46
Dive (20 min, dusk)	70	4.0	3.5, 45°	46
Drift (75 min, night)	46	2.4	4.7, 0°	36
Caged {~19 °C}	19	0.6		19
Day (~22 °C)	32	1.5		29
Night (~22 °C)	40	2.0		33
Overall (24 h, free)	36	1.8	0.8	31

* Estimated at 17 °C.

shown in Fig. 2. Areas under mantle cavity pressure curves (Fig. 2c) were related to oxygen consumption or speed with laboratory calibrations⁶, and changes in baseline (when differential pressure was zero while the mantle refilled) were correlated with depth changes. One of the tracked *Nautilus* was lost shortly after release, the other behaved as previously reported⁵ (Fig. 1). Detailed analysis of short movement sequences was possible because high-resolution fixes (1–2 m) from the Global Positioning System (GPS) were temporarily accessible to the public during and for a short period after the Gulf War in 1991. Long-term behavioural trends can be seen in activity spectra (Fig. 3a), which show how often a pressure occurs in particular data subsets. Average oxygen consumption rates (power) can be calculated by transforming individual pressures to oxygen

Activity levels of *Nautilus* in the wild

R. K. O'Dor*, J. Forsythe†, D. M. Webber*, J. Wells‡ & M. J. Wells‡

* Biology Department, Dalhousie University, Halifax B3H 4J1, Canada

† Marine Biomedical Institute, University of Texas, Galveston, Texas 77555-0843, USA

‡ Zoology Department, Cambridge University, Cambridge CB2 3EJ, UK

WHILE ammonites and all other ectocochleate cephalopods became extinct, nautiloids survived relatively unchanged from the Ordovician, suggesting that they are unusually well adapted to their niche. Here we obtain high-resolution tracks of *Nautilus* positions and depths, combined with telemetered jet pressures, which clarify both its lifestyle and economics. *Nautilus* is more active in nature than in captivity¹, but its energy costs are lower than projected^{2,3}. Viewing *Nautilus* as 'vertic', rather than benthic, resolves this contradiction. Records show that the cost of transport is the same in any direction within a vertical plane. Living on a reef face swept by a lateral current means that vertical movements^{4,5} sample large areas for chemical trails. A detected trail can be followed upcurrent in the slow-moving boundary layer, but no effort is wasted on horizontal movement without good prospects for food; long-range movements are downcurrent and made by drifting. Once fed, a *Nautilus* can reduce its energy costs by moving to deeper, cooler waters, where a single meal can last for months.

Nautilus were trapped and released off the edge of the barrier reef at 225–300 m around a site near the University of Papua New Guinea's Motupore Island Research Station (Fig. 1b). They were fitted with differential pressure transmitters as before⁶ (Vemco, Nova Scotia) to produce pressure records like those

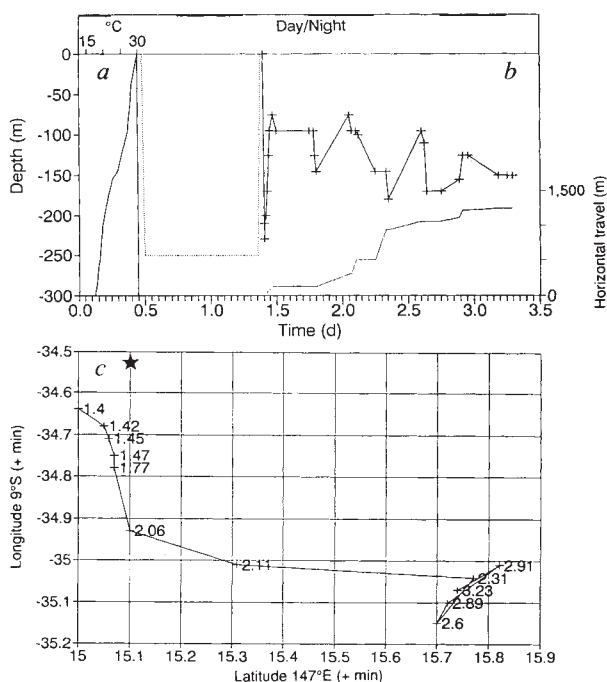


FIG. 1 Three-dimensional positions of the tracked *Nautilus* in relation to the temperature profile (a). Depths are indicated in b by a dashed line during the caged period and solid line when the animal was free (+ marks recording periods). Cumulative horizontal travel (dotted line, bottom right) and speed (slope of line) are also shown in b. Light and dark (thick line) periods are marked at the top; star (in c) represents a marker buoy and trap tie on top of the reef. c, Actual animal positions determined by GPS at times given in decimal days from 00:00 h, 1 March 1991.

consumption and accumulating them as shown in Fig. 3b. Average consumption or speed can be back-calculated from the pressures equivalent to 50 per cent power consumption^{6,7}. Short- and long-term estimates from this approach are shown in Table 1.

Field records comparable to laboratory records were obtained by leaving the *Nautilus* in a cage on the sea floor at 250 m (18 °C) for one day before releasing it with a messenger. The caged *Nautilus* behaved very similarly to laboratory captives^{1,6}, with long quiet periods interspersed with brief periods of activity (Fig. 2a–c). But when freed, the *Nautilus* became much more active. In the period of greatest activity, immediately after release (Fig. 2d), it moved up from 250 m to a position on the reef face at 75 m (almost 27 °C). It then moved down and along the reef face (Fig. 1), covering 1 km in the first 24 h at speeds up to 5 cm s⁻¹ (4.3 km day⁻¹). On the second day it settled down to a more typical pattern of inshore (up-reef) to offshore (down-reef) movement in the day, moving a total of only 0.3 km. As previously reported⁵, activity was predominantly at night (Fig. 3; Table 1) and vertical movements were mostly crepuscular, but not exclusively, up at dusk, down at dawn (Fig. 1a).

Table 1 shows that *Nautilus* is energetically very conservative. Even its maximal exertion was well within its aerobic capacity. Climbing and diving cost about the same as horizontal swimming at the same speed in the laboratory, and the major horizontal displacement costs much less, reflecting assistance by the current. More energy is used at night, but even daytime power

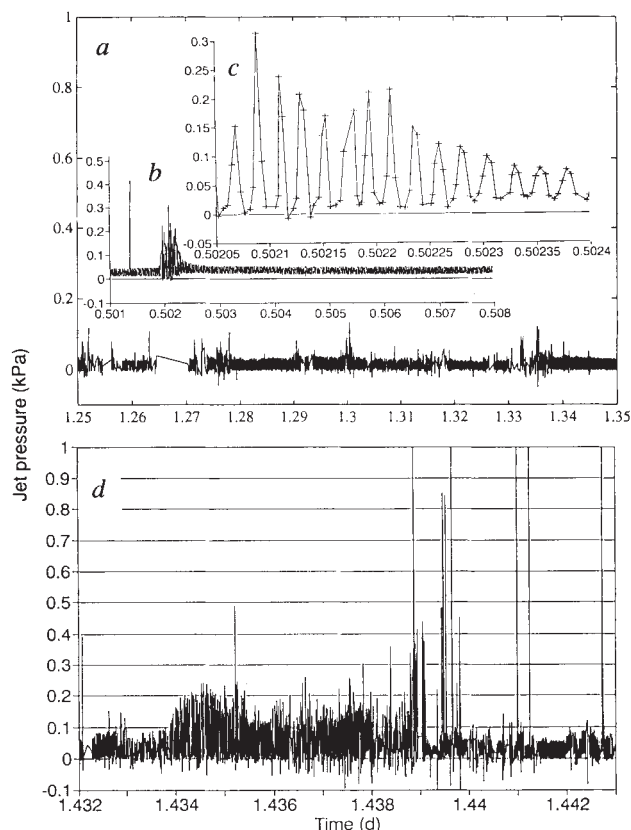


FIG. 2 Mantle cavity pressure records. *a*, A long continuous record shows quiescence with occasional bouts of activity from the second day of the caged period. Insets *b* and *c* show details from the first day under optimal conditions: *b* features a single moderate jet and a bout of spontaneous activity on a background of quiet respiration, and *c* illustrates the detailed form of some of these jet/respiratory cycles. *d*, Period of maximal activity just after the *Nautilus* was released from the cage, as it swam up to rest on the reef. These data have been filtered and the depth component of the baseline extracted. (Note: 0.01 day = 0.25 h, 0.0001 day = 8.6 s).

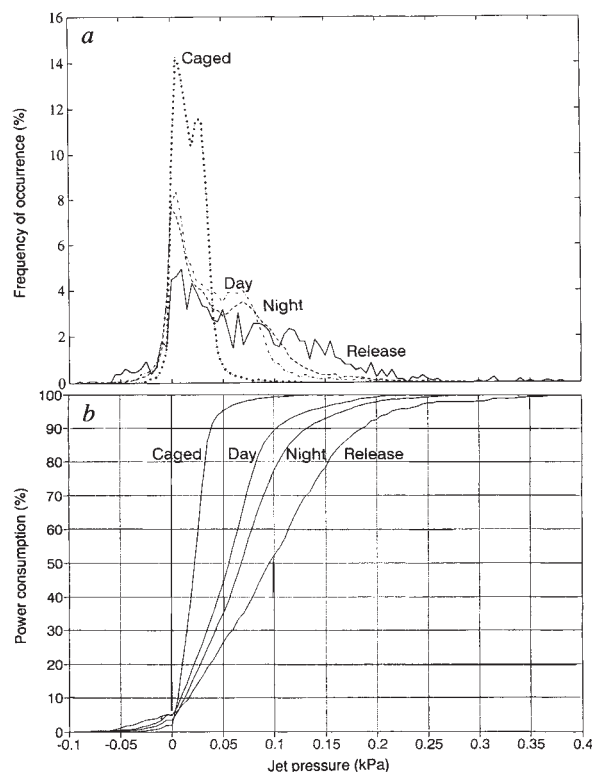


FIG. 3 Activity spectra in *a* indicate the frequency of occurrence of various pressures from the combined records for (1) the caged period, (2) all-night periods (between $x + 0.75$ days and $x + 1.25$ days), (3) all-day periods, and (4) the maximum activity period after release (see Fig. 2d, 1.434–1.439 days). Spectrum (1) is dominated by quiet respiration, and (2) and (3) include pressures in the 0.05–0.1 kPa range associated with 'maximum-economy swimming'⁶. The 'release' spectrum (4) includes pressures up to 0.3 kPa (maximum aerobic pressure) and a few higher pressures only produced by the action of anaerobic head retractor muscles⁶. In *b*, spectra are converted to oxygen consumption ($V_{O_2} = 218p^{0.584}$; V_{O_2} in ml O₂ kg⁻¹ h⁻¹, convertible to power; pressure, p , in kPa) and plotted cumulatively as a percentage of total consumption to illustrate the distribution of power consumption over various levels of activity⁷. Because the original V_{O_2} equation was calculated from average pressures, the pressure at 50% power consumption can be used to back-calculate the average V_{O_2} for spectra using the equation above to produce values as reported in Table 1.

consumption when *Nautilus* is free is higher than in the cage. This reflects the choice of higher temperatures (estimated average, 22 °C), but also that much less time is spent in the quiescent state seen in Fig. 2a and b. As captives were unfed, freedom may simply have provided more stimuli to search for food. Nonetheless, direct conversion of overall pressure to oxygen consumption at 17 °C predicts a daily rate of 744 ml O₂ kg⁻¹, and even at 22 °C this should be less than the 1,190 ml O₂ kg⁻¹ estimated previously^{2,3}. A daily 456 ml O₂ kg⁻¹, as recorded in the cage, equates to only 750 mg of food, so one 50-g cropful would last an unlucky, conservative *Nautilus* at the reef bottom for over two months³.

Unlike most fishes, the locomotor design of *Nautilus* is almost symmetrical in a vertical plane⁸, and there appear to be no extra costs for buoyancy regulation during depth changes⁹. As its cost of transport at low speed is lower than that of fishes, it seems an ideal vertic scavenger. Modern coleoid cephalopods and many zooplankters migrate vertically in marine thermal gradients to 'hunt warm' and 'rest cool'¹⁰ and many benthic scavengers smell currents. But, does any other organism combine these two approaches so efficiently on vertical reef faces; has one ever? Only more attention to vertic ecology can answer this question. □

Received 18 November 1992; accepted 9 February 1993.

1. Zann, L. *Veliger* **27**, 19–28 (1985).
2. Wells, M. J. in *Nautilus, The Biology and Paleobiology of a Living Fossil* (eds Saunders, W. B. & Landman, N. H.) 339–350 (Plenum, New York, 1987).
3. Wells, M. J. *Veliger* **30**, 69–75 (1987).
4. Carlson, B., McKibben, J. & de Gruy, M. *Palau. Pacif. Sci.* **38**, 183–188 (1984).
5. Saunders, W. B. & Ward, P. D. in *Nautilus, The Biology and Paleobiology of a Living Fossil* (eds Saunders, W. B. & Landman, N. H.) 137–162 (Plenum, New York, 1987).
6. O'Dor, R. K., Wells, M. J. & Wells, J. J. *exp. Biol.* **154**, 383–396 (1990).
7. O'Dor, R. K. *et al. Mar. Behav. Physiol.* (in the press).

8. Chamberlain, J. A. Jr in *Nautilus, The Biology and Paleobiology of a Living Fossil* (eds Saunders, W. B. & Landman, N. H.) 489–525 (Plenum, New York, 1987).
9. Ward, P. & Martin, A. W. *J. exp. Zool.* **205**, 5–12 (1978).
10. McLaren, I. A. *J. Fish. Res. Bd Can.* **20**, 685–727 (1963).

ACKNOWLEDGEMENTS. This work was funded by an NSERC (Canada) International Collaborative Grant to R.K.O. and a Royal Society (UK) grant to M.J.W. and was presented at the 29th IGC Congress Symposium on Living and Fossil Cephalopods, Kyoto, Japan (abstracted in *J. Cephalopod Biol.*). We thank the Motupore Island staff, Y. Andrade and S. O'Dor for data reduction, K. Bailey of Furuno for loaning us a GP-1500 GPS Plotter, and the US military for the satellites that made it work.

Body mass dependence of H^+ leak in mitochondria and its relevance to metabolic rate

Richard K. Porter & Martin D. Brand

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK

THE standard metabolic rate of an animal is the rate of heat production under conditions that minimize known extra requirements for energy^{1,2}. In tissues and cells from aerobic organisms, energy expenditure can conveniently be measured as oxygen consumption^{3,4}. Measurements made using isolated rat hepatocytes have shown that a significant contribution to resting oxygen consumption (and hence heat production) is made by a futile cycle of proton pumping and proton leak across the mitochondrial inner membrane⁵. Two important factors affecting standard metabolic rate, thyroid status and phylogeny, also affect the proton permeability^{5–10}. A third major factor affecting standard metabolic rate is body mass^{11,12}. Here we show that proton leak decreases with increasing body mass in mammals. We suggest that differences in proton leak may partly explain the differences in standard metabolic rate between mammals of different mass.

The leaking of protons by diffusion across the mitochondrial inner membrane can explain the less than perfect coupling of

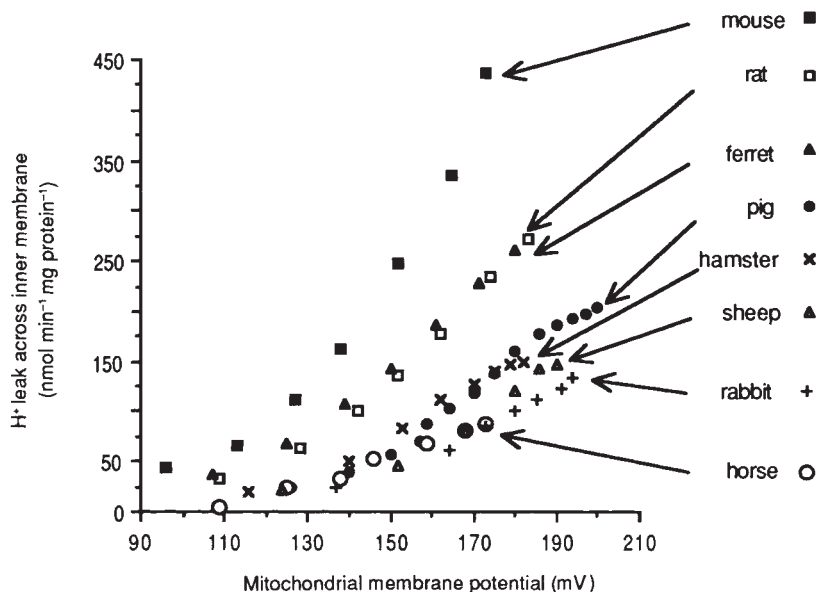
oxygen consumption to ATP synthesis^{13–16}. The kinetics of proton leak are measured by titration of the respiration rate and membrane potential with an inhibitor of the electron transport chain¹⁵. The rate of proton pumping and proton leak is a linear function of respiration rate, assuming that the number of protons pumped per oxygen molecule by the electron transport chain at different potentials is the same^{17–19}. Proton leak is driven by the protonmotive force, which in isolated mitochondria can be expressed as a membrane potential, so that a plot of respiration rate as a function of membrane potential is also a plot of proton leak as a function of driving force, or a current–voltage curve for proton leak across the mitochondrial inner membrane.

Proton leak is a significant contributor to the basal respiration (about 25%) in isolated rat hepatocytes⁵ and so is not just an artefact of isolated mitochondria. It has been proposed that proton leak may be a significant contributor to standard metabolic rate in the whole animal²⁰. Certainly thyroid hormone, a potent inhibitor of standard metabolic rate, has a dramatic stimulatory effect on proton leak in isolated mitochondria^{6–9} and in isolated cells⁵. There are also phylogenetic differences in the amount of proton leak¹⁰: liver mitochondria from a reptile (*Amphibolurus vitticeps*) have greatly reduced proton leak compared to those from a rat, a mammal with the same body size and body temperature; the lizard has a standard metabolic rate that is 20% that of the rat. The third major factor affecting standard metabolic rate, body mass, is considered here.

The relationship between standard metabolic rate (E/t) and body mass (M) in eutherian mammals is given by $E/t = aM^{0.75}$,

FIG. 1 Comparison of proton leak kinetics in liver mitochondria isolated from mammals of different body mass.

METHODS. Mitochondria (3 mg protein), isolated from liver of freshly killed mouse (0.02 kg), hamster (0.1 kg), rat (0.25 kg), ferret (1 kg), rabbit (2.5 kg), sheep (33 kg), pig (50 kg) and horse (150 kg) by standard methods normally used for rat¹⁸, were incubated at 37 °C in 3 ml medium containing 120 mM KCl, 5 mM HEPES–KOH, pH 7.2, 1 mM EGTA, in the presence of 7.5 mM succinate (K^+ salt), 5 μ M rotenone, nigericin (100 pmol per mg protein), oligomycin (1 μ g ml^{−1}) and 5 μ M methyltriphenylphosphonium (TPMP). Respiration rates and membrane potentials were titrated with 0–10 mM malonate (K^+ salt). Each determination was made in either triplicate or quadruplicate. Membrane potentials were measured using a TPMP electrode. Proton leak rate was calculated from respiration rate by assuming 6 protons were pumped (and leaked) per oxygen atom. TPMP binding corrections, mitochondrial volumes and respiration rate were determined as described¹⁰. Error bars are omitted for clarity, but were less than 8% for membrane potential and 19% for proton leak rate in all cases. To determine whether the differences in proton leak per unit mass of mitochondrial protein could be attributed to a greater amount of non-mitochondrial protein present in the mitochondrial preparations from larger mammals compared with those from smaller mammals, horse and mouse liver mitochondrial preparations were assayed for protein and succinate dehydrogenase activity after discontinuous Percoll density gradient centrifugation. More than 90% of the protein present in the horse mitochondrial preparation coincided with the band containing 98% of the succinate dehydrogenase activity. Similarly, more than 60% of the protein present in the mouse mitochondrial preparation coincided with the band



containing 95% of the succinate dehydrogenase activity. It was concluded that greater non-mitochondrial protein contamination in the horse mitochondrial preparation could not account for the differences measured for proton leak rate when compared with the preparation for mouse mitochondria.

TABLE 1 General features and the properties of liver mitochondria of different mammals

Gender	Mouse (<i>Mus musculus</i>) females	Hamster (<i>Merocricetus</i> <i>auratus</i>) females	Rat (<i>Rattus</i> <i>norvegicus</i>) females	Ferret (<i>Mustela furo</i>) males	Rabbit (<i>Oryctolagus</i> <i>cuniculus</i>) mixture	Sheep* (<i>Ovis aries</i>) females	Pig (<i>Sus</i> <i>domesticus</i>) females	Horse (<i>Equus</i> <i>caballus</i>) females
Body mass in kg†	0.02 ± 0.00 (24)	0.1 ± 0.0 (18)	0.25 ± 0.00 (21)	1.0 ± 0.2 (3)	2.5 ± 0.1 (3)	33.0 ± 0.0 (3)	53.0 ± 3.0 (3)	150‡ (3)
Calculated standard metabolic rate (293 M ^{0.75} , kJ day ⁻¹)	16	52	104	293	583	4,034	5,755	12,558
Calculated mass- specific metabolic rate (293 M ^{-0.25} , kJ day ⁻¹ kg ⁻¹)	779	521	414	293	233	122	109	84
Mitochondrial characteristics								
Respiratory control ratio (state III _{uncoupler} / state IV)	2.3 ± 0.4 (3)	4.5 ± 0.7 (3)	3.5 ± 0.1 (19)	3.6 ± 0.8 (3)	5.1 ± 0.4 (3)	9.0 ± 2.0 (3)	9.0 ± 2.0 (3)	7.0 ± 1.0 (3)
State IV matrix volumes (μl per mg protein)	0.9 ± 0.1 (3)	0.9 ± 0.2 (3)	0.97 ± 0.08 (21)	1.2 ± 0.3 (3)	0.7 ± 0.1 (3)	0.8 ± 0.1 (3)	1.1 ± 0.3 (3)	0.9 ± 0.2 (3)
TPMP binding correction (μl per 3 mg protein) ⁻¹	0.13 ± 0.01 (3)	0.27 ± 0.03 (3)	0.19 ± 0.05 (5)	0.17 ± 0.03 (3)	0.33 ± 0.01 (3)	0.36 ± 0.05 (3)	0.20 ± 0.02 (4)	0.20 ± 0.02 (3)

All mammals were fed *ad libitum*. Mice, hamsters and rats were killed by cervical dislocation; all other mammals were injected with a lethal dose of pentobarbitone. Mitochondrial experiments were repeated for *n* separate mitochondrial preparations and each determination was made at least in triplicate (mean ± s.e.m.(*n*)).

* All sheep were lactating.

† Body mass is expressed as mean ± s.e.m. (number of mammals).

‡ Approximate value.

where *a* is the elevation constant and has a value of 293 when units of kJ per day and kg are used^{1,2}. As the mass-specific metabolic rate (metabolic rate per unit mass) has an exponent of -0.25, an increase in body mass of 10,000-fold is associated with a decrease in mass-specific metabolic rate of 10-fold. Differences in mass-specific metabolic rate are due to differences in the amount of metabolically active organs²¹ and in the metabolic rate of different organs^{22,23}. The liver is a significant contributor to basal metabolism, accounting for ~20% of standard metabolic rate in the rat²⁴. The relative rate of oxygen consumption of tissue slices is greater for smaller mammals⁴, and although the biochemical basis of this is unknown, changes in the amount of metabolically active tissue and in the rate of oxygen consump-

tion per unit mass of that tissue are paralleled by changes in the content of mitochondria. The total number of liver mitochondria in a mammal has been shown to be proportional to *M*^{0.72} (ref. 25), and furthermore, when measurements of mitochondrial inner membrane surface areas from tissues of mammals are summated, they show a direct correlation with standard metabolic rate for the whole animal²¹.

We selected mammals ranging in body mass from 20 g (mice) to 150 kg (horses), which represents a 7,500-fold difference in body mass and a roughly 9.3-fold difference in the calculated mass-specific metabolic rate (Table 1). Properties of liver mitochondria isolated from these mammals were measured (Table 1) and the proton leak kinetics determined. Figure 1 shows that the proton leak kinetics are nonlinear and differ for each mammal. In general, there is a decrease in proton leak rate with increasing body mass of mammal. This result was anticipated by Calder's suggestion that "geometry and ionic conductance of membranes had to evolve together"²⁶. We then selected a plausible physiological membrane potential of 170 mV and plotted the log of proton leak rate at that potential for each of the mammals against the log of body mass (Fig. 2). An exponent of -0.14 was determined for this log-log plot.

Hence with an increase in body mass of 10,000-fold, a decrease in mass-specific metabolic rate of 10-fold and a decrease in mitochondrial proton leak per unit mass of mitochondrial protein of 3.6-fold would be expected. This difference is compounded by the fact that larger animals have fewer mitochondria per kg body mass²⁵, and so the oxygen consumption per kg needed to drive the proton leak will be even less in larger mammals. If mitochondrial proton leak is indeed a significant contributor to standard metabolic rate²⁰, then we suggest that changes in this proton leak caused by changes in the intrinsic permeability of the inner membrane and by changes in the amount of mitochondria per kg body mass may explain a significant part of the differences in standard metabolic rate between mammals of different mass. □

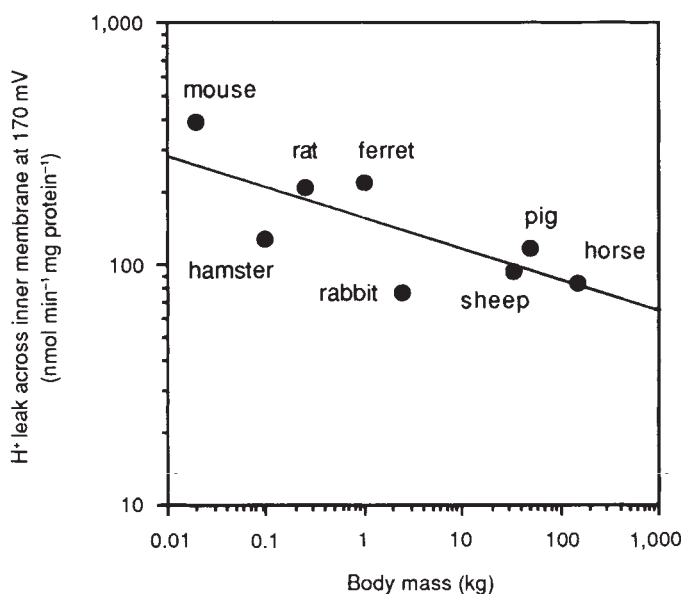


FIG. 2 Relationship between body mass and rate of proton leak at a membrane potential of 170 mV across the liver mitochondrial inner membrane. The line was fitted by linear regression (correlation coefficient, 0.59) and had a slope (exponent) of -0.14.

Received 1 December 1992; accepted 1 February 1993.

1. Kleiber, M. in *The Fire of Life* (Wiley, New York, 1961).

2. Brody, S. in *Bioenergetics and Growth* (Reinhold, New York, 1951).

3. Kleiber, M. *Proc. Soc. exp. Biol. Med.* **48**, 419-423 (1941).
4. Krebs, H. A. *Biochim. biophys. Acta* **4**, 249-269 (1950).
5. Nobes, C. D., Brown, G. C., Olive, P. N. & Brand, M. D. *J. biol. Chem.* **265**, 12903-12909 (1990).
6. Hafner, R. P., Nobes, C. D., McGown, A. D. & Brand, M. D. *Eur. J. Biochem.* **178**, 511-518 (1988).
7. Hafner, R. P., Leake, M. J. & Brand, M. D. *FEBS Lett.* **248**, 175-178 (1989).
8. Horrum, M. A., Tobin, R. B. & Edlund, R. E. *Molec. cell. Endocrinol.* **68**, 137-141 (1990).
9. Brand, M. D., Steverding, D., Kadenbach, B., Stevenson, P. M. & Hafner, R. P. *Eur. J. Biochem.* **206**, 775-781 (1992).
10. Brand, M. D., Couture, P., Else, P. L., Withers, K. W. & Hulbert, A. J. *Biochem. J.* **275**, 81-86 (1991).
11. Kleiber, M. *Hilgardia* **6**, 315-353 (1932).
12. Brody, S. & Procter, R. C. *Univ. Mo. Agr. Exp. Sta. Res. Bull.* **166**, 89-101 (1932).
13. Mitchell, P. & Moyle, J. *Biochem. J.* **105**, 1147-1162 (1967).
14. Krishnamoorthy, G. & Hinkle, P. C. *Biochemistry* **23**, 1640-1645 (1984).
15. Brand, M. D. *Biochim. biophys. Acta* **1018**, 128-133 (1990).
16. Brown, G. C. & Brand, M. D. *Biochim. biophys. Acta* **1059**, 55-62 (1991).

17. Brown, G. C. *J. biol. Chem.* **264**, 14704-14709 (1989).
18. Hafner, R. P. & Brand, M. D. *Biochem. J.* **275**, 75-80 (1991).
19. Brand, M. D. & Dolez, P. *EBEC Short Rep.* **7**, 64 (1992).
20. Brand, M. D. *J. theor. Biol.* **145**, 267-286 (1990).
21. Else, P. L. & Hulbert, A. J. *Am. J. Physiol.* **248**, R415-R421 (1985).
22. Schmidt-Nielsen, K. in *Scaling: Why is Animal Size so Important?* (Cambridge Univ. Press, Cambridge, 1984).
23. Blaxter, K. L. in *Energy Metabolism in Animals and Man* (Cambridge Univ. Press, Cambridge, 1989).
24. Jansky, L. *Acta Univ. Carol. Biol.* **1**, 1-91 (1965).
25. Smith, R. E. *Ann. N.Y. Acad. Sci.* **62**, 403-422 (1955).
26. Calder, W. A. in *Size, Function and Life History* (Harvard Univ. Press, Cambridge, 1984).

ACKNOWLEDGEMENTS. We thank G. Allgood, J. Richardson, J. Vandenburg, A. Grace, A. Fowden, M. Silver, R. Eastwood, L. Hall, G. Cremona and M. Carrington for tissue sources, and M. Leach for technical assistance. This work was funded by a grant from the Agricultural and Food Research Council.

Homeotic genes autonomously specify one aspect of pattern in the *Drosophila* mesoderm

Stephen Greig & Michael Akam

Wellcome/CRC Institute and Department of Genetics, Tennis Court Road, Cambridge CB2 1QR, UK

TRANSPLANTATION and ablation experiments have led to the generalization that in insects the mesoderm is naive, and that pattern is imposed upon it by the ectoderm¹⁻⁶. This has been demonstrated directly by mosaic analysis for the case of one muscle in *Drosophila*. The unique character of this muscle depends on the activity of sex-determining and homeotic genes, not in the muscle itself, but in the nerve that innervates it⁷. Indirect evidence suggests, however, that homeotic genes specify some aspects of mesoderm patterning autonomously. Homeotic genes are expressed in the mesoderm, and are regulated in a segment-specific pattern analogous to, but different from, that seen in the ectoderm^{6,8}. Moreover, the effects of homeotic mutations on the muscles do not always mirror transformations seen in the epidermis^{9,10}. Here we examine this problem directly, by expressing homeotic genes ectopically in the mesoderm without altering their expression in the overlying ectoderm. We find that the pattern of adult muscle precursor cells characteristic of the thorax can be converted to

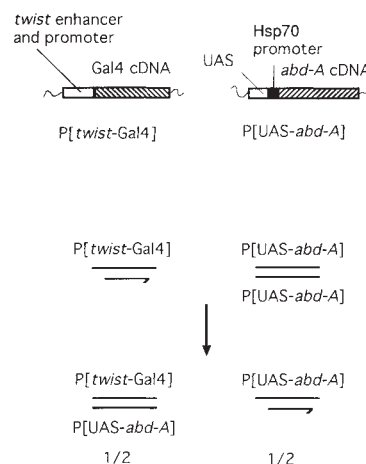
that seen in the abdomen by expressing the homeotic gene *abdominal-A* specifically in the mesoderm.

We expected that mis-expression of homeotic genes would kill embryos, preventing the recovery of transgenic flies carrying constitutively expressed constructs. We therefore used an inducible expression system¹¹. This exploits the ability of the yeast GAL4 protein to activate promoters in *Drosophila*¹² (see also ref. 13 for the application of a similar system in mice). We constructed one transgenic strain in which a mesoderm-specific promoter (from the *twist* gene^{14,15}) was used to direct expression of GAL4 protein. In a second strain, we placed the homeotic gene *abdominal-A* (*abd-A*)^{16,17} under the control of an upstream regulatory element that is activated by GAL4 (Fig. 1). (In normal development, *abdominal-A* specifies the development of abdominal segments.) Embryos from both of these strains develop with no obvious abnormality. But when flies from the two strains are crossed, a fraction of the progeny carry both the *P[*twist-Gal4*]* construct driving GAL4 expression, and the *P[UAS-*abd-A*]* construct responding to GAL4. Such embryos express *abd-A* protein ectopically throughout their mesoderm (Fig. 2). They die at around the time of hatching.

The ectopic expression of *abd-A* is detected at about stage 7 (ref. 18), shortly before endogenous *abd-A* protein is detected. The levels of ectopic *abd-A* protein in the mesoderm appear to be comparable to the endogenous levels in the abdominal mesoderm of normal embryos, until at least stage 11, when they begin to decline in most mesodermal cells. Thus the *abd-A* protein induced indirectly from the *twist* promoter appears

FIG. 1 Scheme for driving *abd-A* expression in the mesoderm. The structures of the two P-element constructs used are shown at the top. Left, *P[*twist-Gal4*]* driver construct; right, *P[UAS-*abd-A*]* responder construct. Boxed areas represent the P-element constructs integrated into the fly genome (thin lines flanking boxes). Enhancer sequences are unshaded; promoter sequences are filled boxes; coding sequences from *Gal4* (left) and *abd-A* (right) are cross-hatched. Bottom, the breeding scheme used to generate the embryos shown in Figs 2 and 3. Males carrying *P[*twist-Gal4*]* inserted on their X chromosome (line 108.4) were crossed to females homozygous for *P[UAS-*abd-A*]*, also on the X (line 21.6). Female progeny carry both *P[UAS-*abd-A*]* and *P[*twist-Gal4*]*, and so are expected to express *abd-A* protein ectopically. Male progeny inherit only *P[*twist-Gal4*]*, and so serve as untransformed controls. In a similar cross, females homozygous for the same *P[*twist-Gal4*]* insert were crossed to a line homozygous for *P[UAS-*abd-A*]* inserted on an autosome (chromosome 2, line 21.9). In this cross all of the progeny inherit both P constructs. All of these embryos show ectopic *abd-A* expression, and the transformed pattern of muscle precursors similar to that depicted in Fig. 3b. A cuticle preparation of a newly hatched larva from this cross is shown in Fig. 3c.

METHODS. Construction of *P[*twist-Gal4*]*: A 3.1-kb fragment of the *twist* enhancer and promoter extending from an upstream *HindIII* site to an artificially created *BamHI* site at the *twist* AUG start codon²¹ was subcloned into the *HindIII* and *BamHI* sites of pBluescript. A *BamHI*-*NotI* fragment containing the *Gal4* cDNA¹¹ was cloned downstream of this using the *BamHI* site at the *twist* AUG and a nearby *NotI* site in the pBluescript polylinker. The *twist-Gal4* sequences were then removed as a *KpnI*-*NotI* fragment and cloned into the corresponding sites of the P-element vector, pCaSpeR 4 (ref. 22). This was then introduced into the genome by P-element-mediated transposition. Construction of *P[UAS-*abd-A*]*: a 1.16-kb *PstI*-*EcoRI* fragment



encompassing the *abd-A* coding sequence, 57 bp of 5' untranslated sequence and 114 bp of 3' untranslated sequence was taken from the *abd-A* cDNA 1dB-5 (ref. 16) and subcloned into the corresponding sites of pBluescript. The fragment was then removed as a *NotI*-*KpnI* fragment and inserted into these sites of *P[UAS]*. *P[UAS]* is a P-element vector with the polylinker downstream of an Hsp70 minimal promoter driven by five copies of the *Gal4* upstream activating sequences (UAS)¹¹.

somewhat later and persists longer than endogenous *twist* protein. This delay presumably reflects the time taken to accumulate effective levels of GAL4 protein in the mesoderm after its transcription is induced, and the time required for these levels to decay after transcription ceases.

The expression of *abd-A* protein in the ectoderm of these embryos is virtually unaffected, as we would expect from the specificity of the *twist* promoter. We see no ectopic protein in the lateral ectoderm (Fig. 2), or in the developing nervous system (data not shown). Ectopic *abd-A* protein is seen in the midline mesectodermal cells, and transiently in a few adjacent ectodermal cells lying 1–2 cell diameters from the ventral midline. This presumably reflects the transient activation of the *twist* promoter in cells lying just beyond the limits of the mesoderm¹⁹.

There are a number of possible markers to assess the patterning of muscle cells. One of the simplest to interpret proves to be the pattern of expression of the *twist* protein itself in late embryos. In the normal embryo, most mesodermal cells cease to express *twist* at around stage 12, when specific mesodermal cell types begin to differentiate. However, the adult muscle progenitor cells continue to express *twist* throughout embryogenesis and larval life²⁰. These cells, analogous to the epidermal cells of the imaginal discs, proliferate during larval development, and differentiate into the specific pattern of adult muscles during metamorphosis. The number and positions of these adult progenitor cells differ between segments—most notably between thoracic and abdominal segments. Abdominal segments A1 to A7 have only six adult muscle precursors in each hemisegment, arranged in a characteristic pattern (Fig. 3a). Thoracic segments

have more adult precursor cells: 15–20, most of which are clustered around the imaginal discs.

We have used the pattern of *twist* expression in stage-14 embryos to assess the pattern of adult muscle precursor cells following the ectopic expression of *abd-A* protein. We find that the thoracic, disc-associated, adult precursor cells are absent. *Twist*-expressing cells are found instead in the positions characteristic for the abdominal segments, A1 to A7. The transformation appears to be almost complete (Fig. 3b; occasionally 1–3 extra *twist*-expressing cells are present in the thoracic segments, in addition to those normally present in abdominal segments). It is not clear whether this transformation of the thoracic to abdominal pattern is achieved solely by suppressing the formation of thorax-specific cells, or whether abdomen-specific cells are generated in the thorax. This is because the normal thoracic pattern includes cells at or close to the positions of those found in the abdominal segments.

In the same embryos, there is no change in the thorax-specific patterns generated by the epidermis. Invaginating imaginal disc primordia are visible in dissected stage-16 embryos stained with toluidine blue (data not shown). The thoracic denticle belts look normal in cuticle preparations of hatching embryos, and Keilin's organs (associated with the imaginal disc rudiments) are present (Fig. 3c). This confirms that no significant levels of *abd-A* protein are present in the epidermis. Thus we conclude that the effect on the pattern of muscle precursors depends autonomously on the homeotic genes expressed in the mesoderm. It remains possible, however, that an inductive signal from the ectoderm is also required to define this pattern correctly.

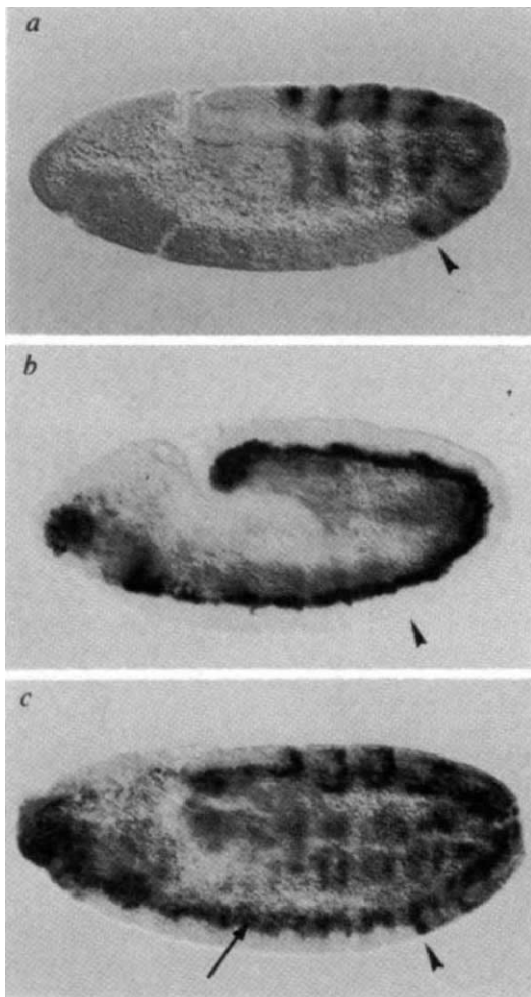


FIG. 2 Ectopic expression of *abd-A* in the mesoderm under the control of the *twist* promoter. *a*, Normal pattern of *abd-A* expression, extending posteriorly from the first abdominal segment (arrowhead marks A1p). *b*, Normal pattern of *twist* expression, limited to the mesoderm (arrowhead marks A1p). *c*, Pattern of *abd-A* expression in embryos carrying one copy of both P[*twist*-Gal4] and P[UAS-*abd-A*]. Note that ectopic *abd-A* expression is visible in the mesoderm of the thoracic segments, but not in the ectoderm (arrowhead marks A1p; arrow marks mesoderm in the thorax). All embryos originated from the same cross (Fig. 1), in which all the progeny are expected to carry P[*twist*-Gal4] but only 50% are expected to carry P[UAS-*abd-A*]. Embryos in *a* and *c* were stained with anti-ABD-A antibody (from J. Casanova) and the embryo in *b* was stained with anti-*Twist* antibody²³. The elite vectastain-ABC kit (Vector) was used according to the manufacturer's instructions. Of 53 anti-ABD-A-stained embryos of stages 9–12, 27 showed ectopic ABD-A expression. All of the germ band extended embryos stained with anti-*Twist* resembled that shown in *b*.

Formation of the final pattern of muscles is not a single event. It depends on a series of patterning processes that take place at different times during development. These integrate both the segment-specific signals provided by homeotic genes, and signals that are common to all segments—provided directly or indirectly by segmentation and dorso/ventral patterning genes. We show that an early cue to define this pattern is provided autonomously by homeotic gene expression in the mesoderm. Recent work (K. Vijayraghavan, J. Fernandes, S. E. Celniker and E. B. Lewis, manuscript submitted) shows that later steps in the development

of the thoracic muscle precursor cells depend on both inductive signals from the epidermis and further signals provided autonomously by homeotic genes expressed in the mesoderm. By allowing the selective and inducible mis-expression of genes, the GAL4 system¹¹ provides a powerful tool to dissect such complex patterning processes. □

Received 11 December 1992; accepted 17 February 1993.

1. Haget, A. *Bull. Biol. Fr. Belg.* **87**, 123–217 (1953).
2. Bock, E. *Roux Arch. dev. Biol.* **141**, 159–247 (1942).
3. Williams, G. J. A. & Caveney, S. *J. Emb. exp. Morph.* **58**, 13–33 (1980).
4. Sahota, T. S. & Beckel, W. E. *Can. J. Zool.* **45**, 421–434 (1967).
5. Beer, J., Technau, G. M. & Campos-Ortega, J. A. *Dev. Biol.* **196**, 222–230 (1987).
6. Bate, M. in *The Development of Drosophila melanogaster* (eds Bate, M. & Martinez-Arias, A.) (Cold Spring Harbor Laboratory, New York, 1993).
7. Lawrence, P. A. & Johnston, P. *Cell* **45**, 505–513 (1986).
8. Akam, M. & Martinez-Arias, A. *EMBO J.* **4**, 1689–1700 (1985).
9. Hooper, J. E. *EMBO J.* **5**, 2321–2329 (1986).
10. Lawrence, P. A. in *Genetics: New Frontiers. Proceedings of the XV International Congress of Genetics* (Oxford and IBH, New Delhi, 1983).
11. Brand, A. & Perrimon, N. *Development* (in the press).
12. Fischer, J. A., Giniger, E., Maniatis, T. & Ptashne, M. *Nature* **332**, 853–856 (1988).
13. Ornitz, D. M., Moreadith, R. W. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **88**, 698–702 (1991).
14. Jiang, J., Kosman, D., Ip, Y. T. & Levine, M. *Genes Dev.* **5**, 1881–1891 (1991).
15. Thisse, C., Perrin-Schmitt, F., Stoetzel, C. & Thisse, B. *Cell* **65**, 1191–1201 (1991).
16. Karch, F., Bender, W. & Weiffenbach, B. *Genes Dev.* **4**, 1573–1588 (1990).
17. Macias, A., Casanova, J. & Morata, G. *Development* **110**, 1197–1207 (1990).
18. Campos-Ortega, J. A. & Hartenstein, V. *The Embryonic Development of Drosophila melanogaster* (Springer, Berlin, 1985).
19. Leptin, M. *Genes Dev.* **5**, 1568–1576 (1991).
20. Bate, M. A., Rushton, E. & Currie, D. A. *Development* **113**, 7–89 (1991).
21. Jiang, J., Hoey, T. & Levine, M. *Genes Dev.* **5**, 265–277 (1991).
22. Pirrotta, V. in *A Survey of Molecular Cloning Vectors and their Uses* (eds Rodriguez, R. I. & Denhardt, D. T.) 437–456 (Butterworth, Boston, 1988).
23. Roth, S., Stein, D. & Nusslein-Volhard, C. *Cell* **59**, 1189–1202 (1989).

ACKNOWLEDGEMENTS. We thank A. Brand and N. Perrimon for communicating their results before publication, for providing us with the P[Gal4] and P[UAS] vectors, and for advice; J. Casanova and S. Roth for antibodies; M. Neale for staining some of the embryos; M. Bate for helping us to interpret them; and M. Bate and P. Lawrence for discussion and for comments on the manuscript. This work was supported by the Wellcome trust.

Clonal dispersion in proliferative layers of developing cerebral cortex

Christopher Walsh*†‡ & Constance L. Cepko*

* Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

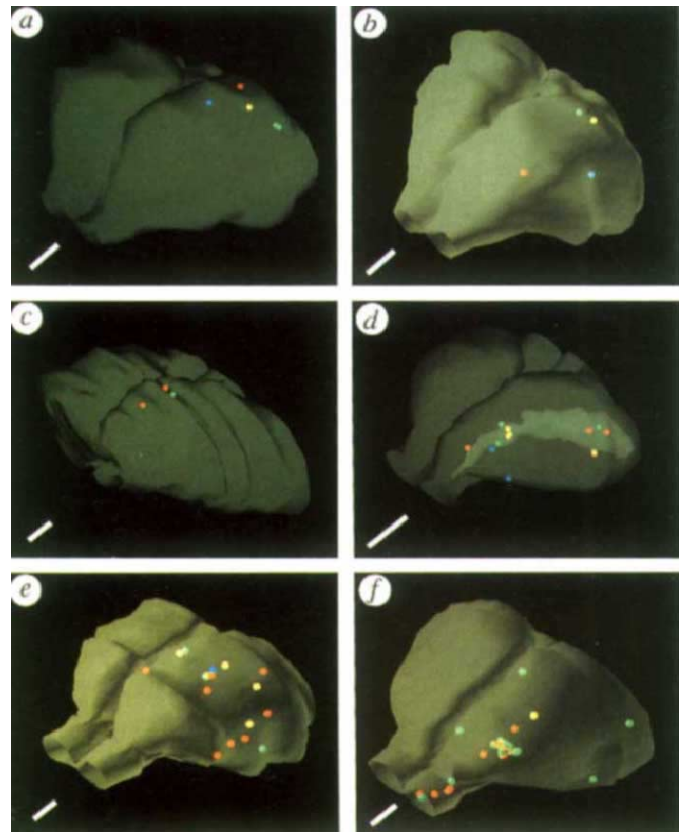
† Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

IN the adult cerebral cortex, many retrovirally labelled clones are widely dispersed¹, though the mechanisms of this dispersion are not well understood. Here we investigate the temporal sequence of clonal dispersion after labelling progenitors of rat cortical cells with replication-incompetent retroviruses at early stages of cortical neurogenesis, 14–15 days after conception (E14/15). The location of labelled daughter cells was determined 3, 6 or 10 days later. Labelled sibling cells were radially arrayed three days after infection (E18). In contrast, by six days after infection (E20/21), 43% of cortical clones were dispersed non-radially by at least 500 μm . Four of these widespread clones were dispersed longitudinally by $\geq 2\text{ mm}$, implying sustained rates of dispersion of $>15\text{ }\mu\text{m}$ per hour. Dispersed sibling cells occurred within proliferative zones of the forebrain in 35% of widely dispersed clones, suggesting that some dispersion reflects movement of dividing cells. Some clones dispersed beyond the neocortex into the olfactory bulb. Progenitor cell dispersion represents a previously unrecognized mode of migration by which sibling cells become widely dispersed in the developing forebrain.

FIG. 3 a and b, Pattern of adult muscle precursor cells in germ band retracted embryos (late stage 14) that are wild-type (a) or express *abd-A* protein throughout the mesoderm (b). Of 47 late stage-14 and stage-15 embryos from the cross shown in Fig. 1, 26 showed the wild-type pattern seen in a. The remaining embryos all had a pattern very similar to that shown in b. The most striking features of this transformed pattern are that the groups of anti-Twist-staining cells found in the wild-type, thoracic segments are missing and the staining pattern in the thoracic segments now strongly resembles that seen in the wild-type abdominal segments A1 to A7. Embryos were stained for *twist* protein as described in Fig. 2. c, Cuticle preparation showing the thoracic segments from an animal similar to b. The thoracic segments are indistinguishable from wild type. Arrow, Keilins organs; arrow-head, ventral pit; T3, denticle belt of thoracic segment 3; A2, denticle belt of abdominal segment 2.

‡ Present address: Department of Neurology, Harvard Medical School, Beth Israel Hospital, 330 Brookline Avenue, Boston, Massachusetts 02215, USA.

FIG. 1 Dispersion of cerebral cortical clones. Three-dimensional computer reconstructions are illustrated from six experiments following injection of the BAG retroviral library at E15. Brains were processed for X-gal histochemistry and PCR analysis of clonal relationships at E18 (*a, b*), E21 (*c, d*) or P3 (*e, f*). Each brain is oriented so that the rostral olfactory bulbs are in the lower-left corner of the picture. Distinct clones are indicated by up to five different colours: red-orange, yellow, green, dark blue and light blue. Each symbol represents a single cell or a tight cluster of cells sharing the same viral tag and are enlarged for clarity (symbol diameter, 120 μm in *a* and *b*, 150 μm in *c* and *d*, 200 μm in *e* and *f*). For each experiment, all histochemically labelled cells in the entire cortex were analysed, but only cells whose DNA tag was successfully amplified are illustrated. Clones consisted of single cells or tight cell clusters at E18. At E21, some clonally related cells were dispersed from one another (such as the red clone in *c*, or red, yellow and green clones in *d*), often parallel to the lateral ventricle (indicated by lighter shading in *d*). At P3, a substantial proportion of clones covered nearly the medial-lateral extent (green clone in *e*), rostral-caudal extent (red clone in *e*), or both dimensions of the cortex (green clone in *f*). Histochemical staining for β -galactosidase showed that widespread clones comprised several cell types, defined using criteria described before^{1,15}. For example, in *f* the green clone included several widely scattered neurons posteriorly, a dense concentration of glial cells in the frontal cortex, and several cells in the subcortical zones and olfactory bulb. Scale bars, 1 mm. **METHODS.** Histochemical techniques were as described¹, except that in order to increase staining sensitivity, some brains processed for X-gal histochemistry at E18 were fixed and sectioned at 60 μm . They were rinsed and reacted for 4 h in 1 mg ml⁻¹ X-gal, 0.5 mg ml⁻¹ nitroblue tetrazolium, in phosphate-buffered saline (pH 7.3) with 2.5 mM MgCl₂. Sections were then rinsed in PBS and coverslipped in Gelvatol. Histochemically labelled cells were plotted using a computer reconstruction package, CARP⁸. One clone in the brain shown in *c*, containing a single cortical cell, was not illustrated.



Retrovirus-mediated transfer of a histochemical reporter gene allows labelling of mitotic cortical progenitor cells and their progeny. Individual clones can be distinguished by using a library of vectors that carry distinct DNA inserts as tags, and by using the polymerase chain reaction (PCR) to analyse the tags¹. The most accurate determination of clonal patterns with this technique requires experiments with few virally labelled clones¹⁻³. The analysis was thus limited to experiments with ≤ 5 PCR-defined clones per cortical hemisphere. In the rat, cortical neurons are generated between days 14 and 21 after conception⁴ (E14–E21, the day of conception being E0), in two proliferative zones: the ventricular zone (VZ) and the subventricular zone (SVZ)⁵. During the period of neurogenesis and the ensuing 7–10 days after birth, post-mitotic neurons migrate through the intermediate zone (IZ) to the cortical plate (CP), the precursor of the cerebral cortex⁶. Therefore labelling cortical progenitors at E14 or 15 should enable most cortical neurons to be analysed.

Cortical clones showed little dispersion when analysed three days after infection at E18 (Fig. 1*a, b*). Clones consisted of single cells (4 clones) or clusters of ≤ 5 cells in the VZ, SVZ and IZ (8 clones). Most of these cells were so closely spaced that multiple cells were processed for PCR in single tissue fragments, making quantitative analysis of clone size difficult (Table 1). As previously noted in studies using a single retrovirus in rat⁷ and mouse^{8,9}, cortical clusters were often oriented radially (Fig. 2*a–c*) as arrays up to 200 μm tall and $\leq 120 \mu\text{m}$ (usually 30–50) in the medial-lateral or rostral-caudal dimension. In one of 11 clusters analysed at E18, PCR analysis showed that nearby cells contained different viral tags, and thus constituted two adjacent clones. All other clusters contained cells with only one viral tag, however, and no widespread dispersion of cells with the same viral tag was seen after analysis of 12 clones at E18 (Table 1).

In contrast to E18 brains, brains analysed at E20/21 (six days after infection) showed very different patterns. Although there were equivalent numbers of viral tags per cortical hemisphere

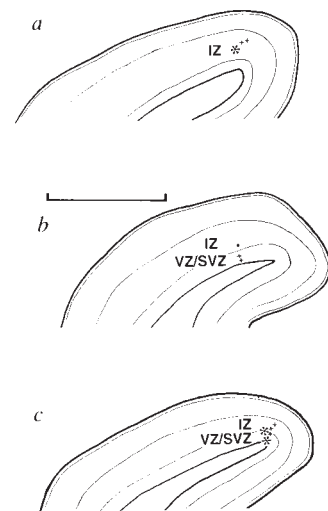


FIG. 2 Laminar locations of retrovirally labelled cells, analysed at E18. *a, b* and *c*, Camera lucida drawings of coronal sections from brains injected with the BAG library at E15 and analysed at E18. In each drawing, dorsal tip is up, with the brain outline shown as a heavy trace. Layers of the developing cortex are indicated with finer lines, and layers containing labelled cells are indicated. Histochemically labelled cells were tightly clustered in the VZ, SVZ and IZ, so that each cluster was included in 1–2 coronal sections taken at 60 μm . PCR analysis showed that cells in 11/12 of these clusters contained one viral tag. Asterisks indicate single cells for which PCR was used to amplify a single viral tag. Dots indicate cells for which PCR amplification was unsuccessful. Such PCR-negative cells are not illustrated in Fig. 1. As labelled cells were dissected for PCR analysis in tissue chunks $\sim 300 \mu\text{m}$ square, two or more tightly clustered cells were sometimes contained in one tissue fragment. Crosses indicate 2–3 cells contained in a single tissue fragment from which a single PCR product was amplified (Table 1). Tight clusters like these are indicated by single dots in Fig. 1*a* and *b*. Scale bar, 1 mm.

TABLE 1 Summary of clonal patterns in the developing cortex

Cortical hemisphere	Clone	Tissue fragments (cells)					Dispersion (mm)	
		Total	VZ/SVZ	IZ/WM	CP	Non-cortex	Longitudinal	Transverse
E18-K, left	A	1 (4)	1 (4)					
	B	2 (3)		2 (3)			0.0	0.04
E18-U, left	A	1			1			
E18-U, right	A	3 (5)	2 (4)	1			0.06	0.05
E18-J, left	A	1 (3)	1 (2)	(1)				
	B	1	1					
	C	2	1	1			0.06	0.1
	D	1 (2)	1 (2)					
E18-J, right	A	1 (3)	1	(2)				
	B	1	1					
	C	1 (2)		1 (2)				
	D	1	1					
E18 total		16 (28)	10 (17)	5 (10)	1			
Average		1.3 (2.3)	63%	31%	6%		0.04	0.06
E21-A, left	A	1			1			
	B	2 (4)			2 (4)		0.0	0.3
	C	1			1			
	D	2	1		1		1.3	~1.4
E21-A, right	A	5 (10)	2	1 (3)	2 (5)		0.2	0.7
E21-H, left	A	3 (5)			3 (5)		0.1	0.3
	A	4	1	1	2		3.1	~0.8
	B	3	2			1 {s/ic}	2.3	~0.3
	C	2		1	1		0.0	0.6
E21-I, left	D	3	1	2			2.3	~0.7
	A	1			1			
	B	5 (7)	4 (6)		1		0.4	0.3
	C	2			2		0.2	0.2
E21-I, right	D	2			2		0.8	~0.4
	A	2	1	1			0.5	0.5
	B	2		2			0.1	0.08
	A*	3	2			1 {s/ic}	2.9	~0.7
E21-C, left	A	1			1			
	B	2			2		1.1	~0.9
	C*	7			1	6 {s/ic}	2.2	~0.8
E20-B, right	A	3 (4)	3 (4)	1			0.1	0.1
	B	1						
E20/21 total		57 (6.9)	17 (20)	9 (11)	23 (30)			
Average		2.7 (3.3)	35%	18%	47%		1.0	0.5
P3-L, left	A	2			2		0.1	3.2
	B	5	1		4		1.0	2.6
	C	1			1			
	D	8		1	7		3.2	3.2
	E	1	1					
P3-M, right	A	1 (2)		1 (2)				
	B	7 (24)	1 (2)			6 (22) {ob}	1.9	0.2
	C	15 (22)	15 (22)				3.1	1.1
P3-S, left	A	21 (26)	2	4 (5)	9 (13)	6 {ob}	3.6	1.3
	B	20 (27)		4 (5)	14 (20)	2 {ob}	6.4	2.6
	C	1			1			
P3-N, right	A	1 (2)		1 (2)				
P3 total		83 (121)	20 (28)	11 (15)	38 (48)			
Average		6.9 (10.1)	29%	16%	54%		2.8	2.0

Results are shown from 18 experiments (defined as a hemisphere of an infected brain) after injection of the BAG retrovirus library at E15 and analysis 3 days later (E18), injection at E14 or E15 and analysis 6 days later (E20/21), or injection at E15 and analysis 10 days later (P3). One clone (asterisk) dispersed across the midline into both hemispheres. Each clone is listed as a row of the table and indicated by A, B and so on. The total number of tissue fragments that contained the same viral tag is indicated for each clone. In case where tissue fragments contained >1 histochemically labelled cell, the total number of labelled cells in all tissue fragments is indicated in parentheses. For each clone, the location of PCR-positive cells in the ventricular or subventricular zones (VZ/SVZ), intermediate zone or subcortical white matter (IZ/WM) or cortical plate (CP) is listed; again, the number of histochemically labelled cells in each fragment is given in parentheses. Non-cortical cells were found that had the same viral tag as cortical cells. They were located in the striatum/internal capsule and in the olfactory bulb (indicated as {s/ic} or {ob}); they were included in calculations of clonal dispersion. Between E18-E21, the cortex increases in length from 3.3 to 5 mm while the average longitudinal dispersion increases more than 20-fold. Between E21 and P3, the cortex grows from 5 to 7.5 mm in length, and longitudinal dispersion increases more than 2.5-fold. With increasing age, clones increase in size, and cells migrate from the VZ/SVZ to the CP. Construction, characterization, and injection of the retroviral library are described elsewhere¹. The library consists of a mixture of retroviral vectors that encode lacZ. Each member of the library also contains a DNA segment of unique size and/or pattern of restriction enzyme digestion. Viral inoculations were made such that ≤5 clones were recovered from each hemisphere. After histochemical processing of tissue sections, labelled cells were removed in small tissue fragments (containing up to ~10,000 unlabelled cells) and digested in a proteinase K solution; viral tags were amplified with PCR. Standard dideoxy sequencing methods were used to determine whether assignment of clonal relationships based upon restriction enzyme digestion patterns was accurate. Sequencing was attempted on 23 amplified tags and was successful on 21 (the other two sequencing reactions yielded no product); DNA sequence analysis showed that the 21 tags had been assigned correctly to 9 clones by the restriction enzyme digestion patterns (data not shown). In 291 clones analysed so far, 85 members of the library have been recovered, suggesting that some members of the library may be absent or present in relatively low amounts². To confirm that differences in clonal patterns were not due to coincidental infections of two or more clones with one viral tag³, the Wilcoxon rank sum test was applied to compare frequencies of dispersed clones at E18 versus E21 and P3. Even though the number of clones per experiment was similar at all ages, the difference in frequency of dispersed clones was significant ($P < 0.0002$). The probability of coincidental double infections can also be calculated directly using statistical models^{2,3} available over e-mail (at church@rascal.med.harvard.edu): When ≤5 clones are labelled in an experiment, the probability that two clones coincidentally carry the same tag is <0.10-0.15, whereas the probability that three or more clones carry the same tag is <0.005; when ≤3 clones are labelled in an experiment, the probability that two clones carry the same tag is <0.05. The longitudinal dispersion separating the most widespread members of each clone was measured by multiplying the number of coronal sections separating the cells by the section thickness. Transverse dispersion was measured by superimposing tracings of cell position when clonally related cells were ≤0.5 mm apart longitudinally. When clonally related cells were dispersed longitudinally >0.5 mm, transverse dispersion was estimated by comparing the medial-lateral position of labelled sibling cells relative to the midpoint of the lateral ventricle. Data from one of these 18 experiments were also included in a previous analysis¹.

(average, 2.4 tags per hemisphere at E18; 2.3 tags per hemisphere at E20/21), cells bearing the same tag were frequently widely dispersed when analysed at E20/21. Forty-three per cent of the labelled clones (9 of 21) were dispersed over $>500\ \mu\text{m}$ (Table 1), and four of these clones were dispersed longitudinally over $>2\ \text{mm}$, a distance equivalent to about half the length of the forebrain at this age (Fig. 1c, d). Dispersion continued postnatally: in clones analysed three days after birth (P3, counting the day of birth as P0), more than half of the labelled clones were widely dispersed (Table 1) and dispersion ranged up to 6 mm (Fig. 1e, f).

Further examination of labelled clones gave an indication of the routes taken by cells during rapid dispersion. Six clones (4 clones analysed at P3, two at E21) contained cells at multiple sites within the telencephalic VZ or SVZ, separated by up to 3.6 mm (average, 2.2 mm). Three of these clones contained cells in the cerebral cortex (Fig. 3a; Table 1); three others showed cells in proliferative regions of the cortex and olfactory bulb, where neurogenesis continues postnatally (Fig. 1f)¹⁰. These patterns suggest that substantial dispersion can occur before cortical cells exit the proliferative layers, as has been clearly illustrated *in vitro*¹¹. Other dispersed clones contained cells in one site of the VZ or SVZ, with other cells in the IZ or CP (five clones; see Fig. 3b); five other clones were restricted to the CP or IZ (Fig. 3b).

Clonal dispersion longitudinally was on average twice as great as transverse dispersion at E21, which largely reflects 4 clones (19%) that were extremely widely dispersed longitudinally. Clones with large longitudinal dispersion were often oriented strikingly parallel to the lateral ventricle (Fig. 1d), perhaps reflecting rapid longitudinal dispersion in proliferative zones (Fig. 3a). Other longitudinal routes of migration, such as longitudinal migration along subcortical fibre tracts, may also occur, however.

Direct observation of migrating cells in ferret cortical slices¹² shows that some ($\sim 10\%$) cells migrate transversely within the IZ (see also refs 4, 13, 14). In our material, clones dispersed in the transverse (medial-lateral) plane as well as longitudinally. Although transverse dispersion was seen prenatally, it was more extensive postnatally (Fig. 1e, f), when many clones were dispersed in both transverse and longitudinal directions. Some of the transverse dispersion seemed to occur after cells left the proliferative zones and moved through the IZ, because it was more extensive after cortical neurogenesis was complete and it involved many cells in the CP (Fig. 3c; Table 1).

Transverse and longitudinal clonal dispersion can also be seen in animals analysed as adults, when most retrovirally labelled clones contain one or more neurons^{1,15}. The decrease in frequency of widespread clones in adults (58% of all clones at P3; 24% in adults) may reflect lower efficiency of PCR analysis (inactivation of the histochemical marker, DNA degradation, poorer tissue lysis). Moreover cell death in the cerebral cortex, which occurs postnatally in rodents¹⁶, may cause this decrease. Cell death, either random or selective for cells in widespread clones, would leave single cells or single clusters remaining from initially widespread clones.

The migration patterns shown by murine cortical cells 2–6 days after retroviral infection has been studied using a single lacZ-encoding retrovirus⁸. Cells were grouped according to geometric criteria, and statistical analysis indicated that most groups were probably clones. To compare patterns in rat and mouse, and to re-evaluate the criteria used to group murine cells, the same grouping criteria were applied to the present results using rat brains. At six days post-infection (E20/21), both the lumping of unrelated cells into groups and the splitting of sibling cells into different groups were very frequent. But dispersion patterns appear to be similar in the two species at shorter survival times, because the same geometric criteria accurately defined clones in rat brains analysed at E18.

As the retroviral library method detects the relative dispersion

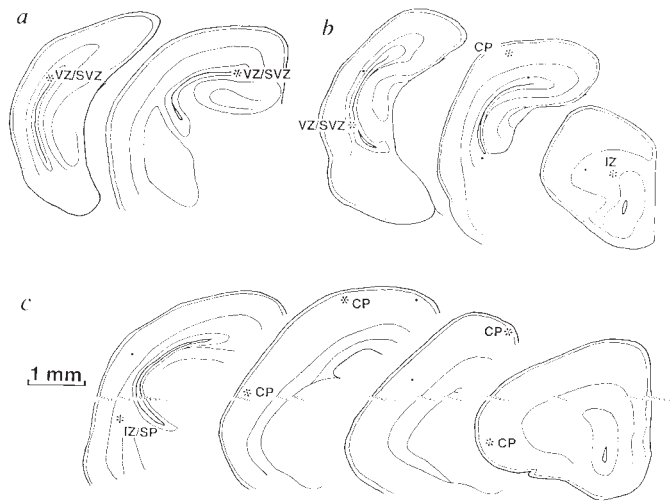


FIG. 3 Laminar locations of retrovirally labelled clones of cells analysed at E21 (a, b) or P3 (c). Asterisks indicate cells sharing the same viral tag. PCR analysis was successful in 40–70% of histochemically labelled cells; cells for which PCR was unsuccessful, or which were part of other clones, are indicated by dots. In a, two VZ/SVZ cells were separated longitudinally by 1.5 mm. In b, labelled cells were dispersed longitudinally by 3.1 mm. Cells of this clone (shown in red in Fig. 1d) occupied the CP and VZ/SVZ, and one cell was in either the IZ or the cortical subplate (SP)¹⁷. Note that it is difficult to calculate precisely the degree to which cells in this clone were dispersed transversely. At P3 (c), cells in the indicated clone (shown in red in Fig. 1e) were located in the CP, except one cell that was either in the IZ or SP, as indicated. Cells in this clone were also dispersed both transversely and longitudinally. Only some of the cells in each clone are illustrated owing to space constraints.

of sibling cells, not the movement of a single labelled cell, it cannot determine the precise age at which non-radial migrations begin. Cell dispersion may occur throughout neurogenesis. Widespread clones may only be detectable with the retroviral technique after several cell divisions of a moving progenitor cell have produced sibling cells at multiple locations. In this case, maximal rates of dispersion would be $\sim 15\ \mu\text{m}$ per hour (if cells dispersed $\geq 2\ \text{mm}$ in 6 days). An alternative possibility is that cortical progenitor cells change their behaviour during development, so that most non-radial dispersion occurs in the later stages of neurogenesis at correspondingly faster rates (maximal rate $\geq 30\ \mu\text{m}$ per hour). Although the roles and mechanisms of clonal dispersion are still being established, our results indicate that several cell types and stages of cortical development involve non-radial movements. □

Received 2 September 1992; accepted 3 February 1993.

- Walsh, C. & Cepko, C. L. *Science* **255**, 434–440 (1992).
- Kirkwood, T. B. L., Price, J. & Grove, E. A. *Science* **258**, 317 (1992).
- Walsh, C. et al. *Science* **258**, 317–320 (1992).
- Altman, J. & Bayer, S. A. *Neocort. Dev.* (Raven, New York, 1991).
- The Boulder Committee. *Anat. Rec.* **166**, 257–262 (1970).
- Rakic, P. *Science* **241**, 170–176 (1988).
- Walsh, C. & Cepko, C. L. *Science* **241**, 1342–1345 (1988).
- Austin, C. & Cepko, C. L. *Development* **110**, 713–732 (1990).
- Luskin, M. B., Pearlman, A. & Sanes, J. R. *Neuron* **1**, 635–647 (1988).
- Bayer, S. A. *Expl Brain Res.* **50**, 329–340 (1983).
- Fishell, G., Mason, C. A. & Hatten, M. E. *Nature* **362**, 636–638 (1993).
- O'Rourke, N. et al. *Science* **258**, 299–302 (1992).
- Rakic, P. *Trends Neurosci.* **4**, 184–187 (1981).
- Misson, J.-P. et al. *Cerebr. Cort.* **1**, 221–231 (1991).
- Grove, E. A., Kirkwood, T. B. L. & Price, J. *Neuron* **8**, 217–229 (1992).
- Finlay, B. L. & Slattery, M. *Science* **219**, 1349–1351 (1983).
- McConnell, S. K., Ghosh, A. & Shatz, C. J. *Science* **245**, 978–982 (1989).

ACKNOWLEDGEMENTS. We thank G. M. Church and E. F. Ryder for help with statistical analysis; C. P. Austin for help with evaluating geometrical clonal criteria; S. K. McConnell, G. Fishell, C. A. Mason and M. E. Hatten for communication of unpublished results; and D. Altshuler, M. H. Chen, D. M. Fekete, E. F. Ryder, C. J. Tabin and J. LoTurco for comments on the manuscript. This research was supported by a grant from the NINDS (C.L.C.) and by a Physician Postdoctoral Fellowship from the Howard Hughes Medical Institute (C.W.).

Dispersion of neural progenitors within the germinal zones of the forebrain

Gord Fishell*, Carol A. Mason & Mary E. Hatten*

Department of Pathology, Center for Neurobiology and Behavior, College of Physicians and Surgeons, Columbia University, 630 West 168th Street, New York, New York 10032, USA

* Present address: The Rockefeller University, 1230 York Avenue, New York, New York 10021-6399, USA

ONE of the early events in the establishment of regional diversity in brain is the subdivision of the forebrain into the cerebral cortex¹⁻⁷ and underlying basal ganglia⁸. This subdivision is of special interest, owing to the striking difference in cellular patterning in these two regions. Whereas the dorsal aspect of the telencephalon gives rise to the laminar, cortical regions of brain, the basal aspect gives rise to nuclear, subcortical regions. To examine early events in the regionalization of the forebrain, we visualized cell movement within the ventricular zones of the dorsal and basal regions of the E15 murine telencephalon. Over an 8–24-hour observation period, labelled cells moved extensively in the plane of the cortical ventricular zone. Cell dispersion was restricted, however, at the border between the cortical ventricular zone and the lateral ganglionic eminence, the basal telencephalic ventricular zone. We suggest that this restriction of cell movements establishes a regional pattern of neurogenesis in the developing brain.

To visualize living cells in the ventricular zone (VZ)⁷ of the developing murine forebrain, the lateral wall of the telencephalic vesicle was removed on embryonic day 15 (Fig. 1) and cells were labelled with the lipophilic dye DiI⁹. Confocal imaging demonstrated that DiI-labelled cells along the ventricular surface had the small size and bipolar morphology characteristic of neural precursors in the VZ¹⁻⁶ (Fig. 2a), expressed cellular antigen markers for VZ cells, including nestin¹⁰ and RC2

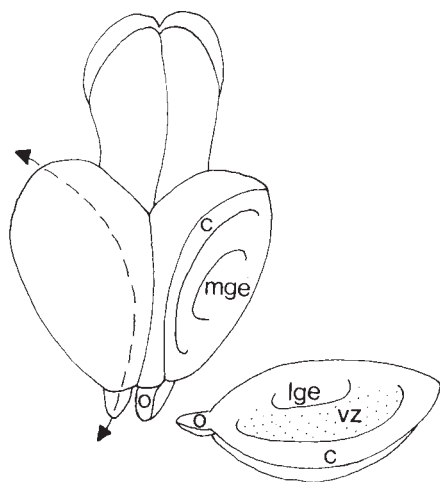
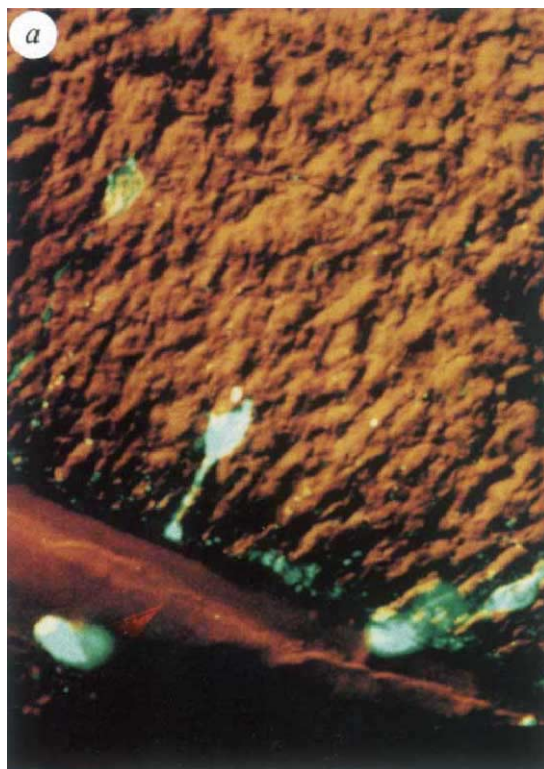


FIG. 1 Preparation of explants of the lateral wall of the telencephalon. Explants of the lateral wall of the E15 mouse telencephalon were prepared by making an incision across the dorsomedial surface of the brain, indicated by the dotted line drawn across the left side of the illustration. The wall of the telencephalic vesicle was subsequently removed by an incision between the lateral and medial ganglionic eminence. The exposed ventricular surface of the explant has a roughly semi-circular shape with a raised central region (the lateral ganglionic eminence) surrounded by a flatter outer region (the cortical ventricular zone). DiI-labelled cells were imaged in the plane indicated by the stippling. The olfactory bulb provided a rostral landmark for the explant. Cortex, C; lateral ganglionic eminence, lge; medial ganglionic eminence, mge; olfactory bulb, o; ventricular zone, vz.



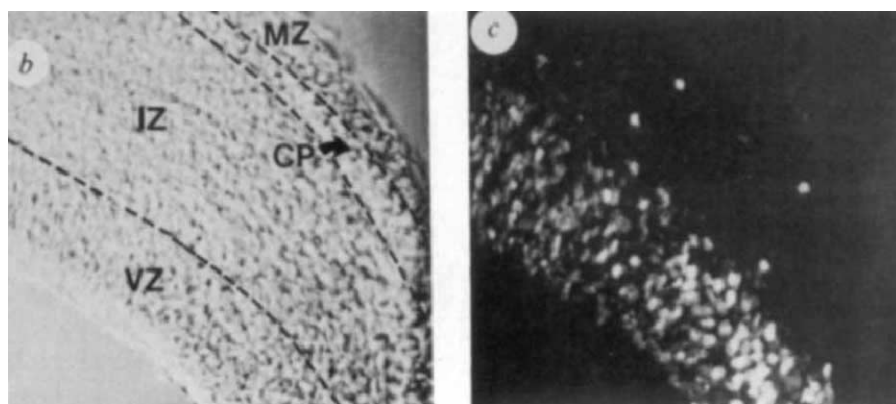
(ref. 11) (not shown), and incorporated bromodeoxyuridine (BrdU) (Fig. 2b, c). This suggests that the explants maintained the pattern of cellular organization and proliferation seen *in situ*^{2,3,7}.

To examine the dispersion of neural progenitor cells in the dorsal and basal telencephalic VZs, DiI-labelled cells were imaged by time-lapse video fluorescence microscopy^{12,13}. Within the dorsal (cortical) telencephalic VZ, labelled cells moved in short bursts, reaching rates of 10–100 μm per hour (Figs 3, 4). Motile cells rapidly extended and retracted thin lamellar processes, 5–20 μm in length, after which the cell body moved in the direction of lamellar extension. Within the basal telencephalic VZ, the lateral ganglionic eminence⁸, cells moved at comparable rates and frequencies (not shown). Thus neural progenitor cells disperse laterally within the germinal zones of the forebrain.

To evaluate whether cells moved within the cortical VZ in a 'random walk', the mean-square displacement of 21 DiI-labelled cells was plotted as a function of time¹⁴. The linearity of these plots indicated that the cells moved in a diffusion mode. The duration of cell movement near the ventricular surface was quite variable, with some cells moving for 8–24 h and other cells exiting the plane of focus soon after cellular division. It was not possible to determine whether the cells that did leave the plane of focus became post-mitotic and attached to radial glial cells, or moved to the apical VZ to re-enter S-phase of the cell cycle. In a few cases ($n = 7$), we imaged labelled cells undergoing cell division. The labelled daughter cells appeared to move independently, taking different trajectories within the VZ and exiting to deeper zones out of the plane of focus at different times.

The dispersion of neural progenitor cells within the telencephalic VZ, as seen here, is consistent with recent reports on the dispersion of retrovirally labelled cells in the ventricular zone of the rat cortex¹⁵. Although initial retroviral marking experiments provided evidence for clustering^{16,17} of clonally related cells in cortex, recently it has been shown that some clonally related cells disperse widely during cortical

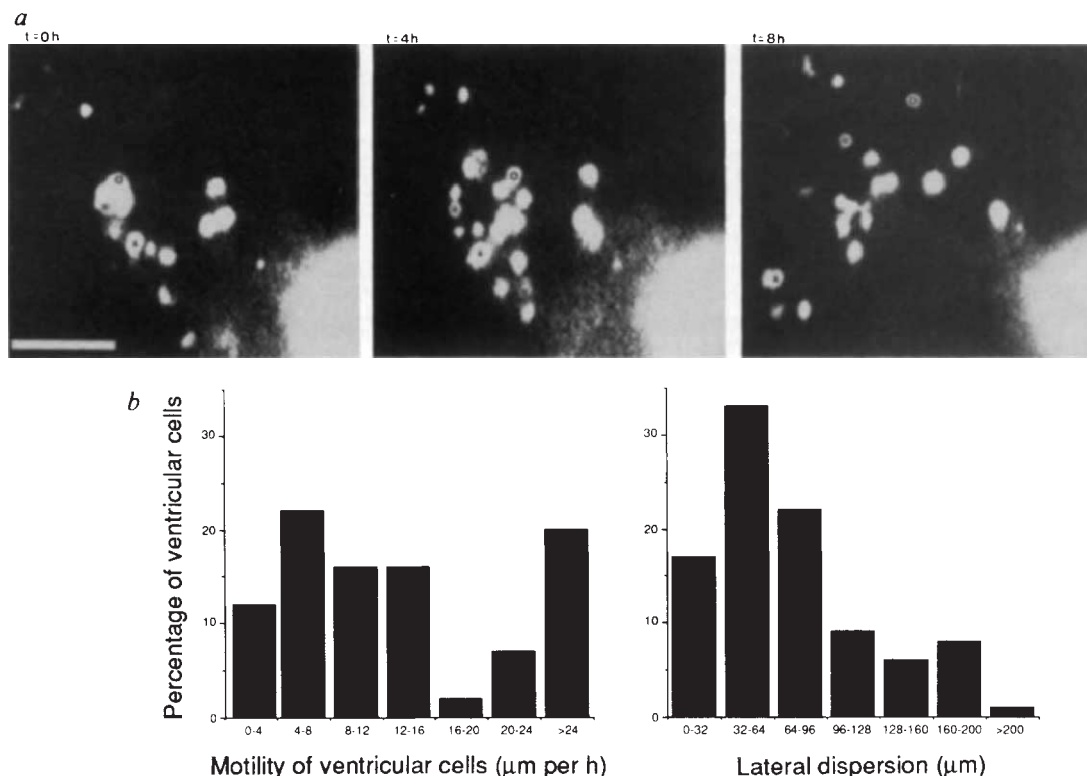
FIG. 2 Dil-labelled cells and BrdU incorporation in E15 cortex explants. To examine the morphology of Dil-labelled cells and determine whether cells continued to undergo cell division *in vitro*, we imaged labelled cells and analysed the pattern of BrdU incorporation in coronal sections of the explants. *a*, Optical sectioning by confocal imaging of Dil-labelled cells within an explant of E15 mouse telencephalon shows labelled cells in the ventricular zone. Ventricular cells can be identified by their position, small size and characteristic, bipolar, apical and ventricular projecting processes. Scale bar, 25 μ m. *b*, By differential interference contrast (DIC) imaging, the VZ, intermediate zone (IZ), cortical plate (CP) and marginal zone (MZ) of the explanted tissue are evident. *c*, BrdU labelling of the explant for 8 h reveals a dense band of proliferating cells in the VZ. Scale bar, 25 μ m.



METHODS. The E15 mouse telencephalic explant was washed in CMF-PBS (4 °C), transferred to DMEM:F12 medium (supplemented with 10% horse serum and 5% fetal calf serum) and incubated in culture medium containing (0.01–0.02%) Dil (1,1'-dioctadecyl 3,3',3'-tetramethyl indocarbocyanine perchlorate; Molecular Probes) at 35 °C for 35'. After washing with fresh medium (35 °C with 5% CO₂), the explant was inverted on a glass coverslip microculture¹² with the ventricular surface facing downwards, overlaid with collagen for 30 min at 35 °C and incubated in fresh medium (35 °C, 5% CO₂) for 8 h. The explant was placed on the preheated (35 °C) stage of a Zeiss Axiovert 35 microscope and time-lapse fluorescent images were acquired as described for Fig. 3. After video imaging, the explant was washed briefly in ice-cold CMF-PBS, fixed in paraformaldehyde (4% in PBS for 4 h), washed 3 times in PBS, removed from the collagen, embedded in agarose and sectioned in the coronal plane with a vibratome. Cells were imaged with a Zeiss Axiovert microscope fitted with Zeiss Plan-Neofluor 20 $\times$ /0.5 or Zeiss Plan-Neofluor 40 $\times$ /0.85 infinity-corrected objectives and a Bio-Rad MRC-600 scan head with 25 mW Argon ion multi-line laser. A high-sensitivity,

green excitation, single-channel filter set was used to acquire images of Dil-labelled cells. A Nomarski image (*b*) was obtained by transmitting the laser light through an optic fibre cable to the scan head before computer acquisition. For BrdU labelling (*c*), BrdU (0.01 mM; 1:100; Amersham, cell proliferation kit) was added to the culture medium at the start of video microscopic observation. To localize BrdU-labelled cells, tissue was treated with 2 M HCl at 50 °C for 1 h, neutralized by washing in PBS (10 $\times$ and 1 $\times$), treated with 10% normal goat serum (30 min) and stained with anti-BrdU antibody (Becton Dickinson; 1:100 in PBS containing 0.02% Triton-X100) overnight (4 °C). After staining, sections were washed in PBS and incubated in fluorescein isothiocyanate-conjugated goat anti-mouse secondary antibody (1:100 for 1 h). Anti-BrdU positive cells (*c*) were imaged with a Zeiss Axiovert microscope fitted with Zeiss Plan-Neofluor 20 $\times$ /0.5 or Zeiss Plan-Neofluor 40 $\times$ /0.85 infinity-corrected objectives, using a Bio-Rad MRC-600 scan head with 25 mW Argon multiline ion laser with blue excitation filter set (BHS).

FIG. 3 The motility and extent of lateral dispersion of cortical ventricular cells. *a*, Time-lapse imaging reveals that ventricular cells disperse in all directions across the cortical VZ. After 8 h, the tight cluster of labelled cells has dispersed over a 200 μ m² area. Positions of three labelled cells (*, $\circ$, Δ) are shown at 4-h intervals. The Dil crystal in the right of the photomicrograph provides a landmark. The tissue is oriented with rostral, top; caudal, bottom; ventro-lateral, left; dorsomedial, right. *b*, Cells move at speeds between 10 and 100 μ m per h and disperse up to 200 μ m over an 8-h observation period. Left, the speed of movement of 100 Dil-labelled cells over a 1-h period indicates varying rates of cell movement. Given that individual cells move in sudden bursts, speeds measured over a 1-h period underestimate the actual speed of motile cells; right, dispersion of the same population of 100 Dil-labelled cells over an 8-h period.



ports for video imaging. Images were acquired with a Hamamatsu intensified charge-coupled device (I-CCD) 2400-50/80 camera head and controller, using a Panasonic time-date generator, Uniblitz shutter and D122 shutter drivers with external source signal input controlled by an Image-1 Image Analysis System (Universal Imaging). Epifluorescent excitation/emission was at 510/550 nm. Neutral density filters were added to reduce the excitation light level by 99%. For time-lapse recordings, one image (100-ms exposure) was acquired every 5 min. No cellular phototoxicity was evident for up to 24 h.

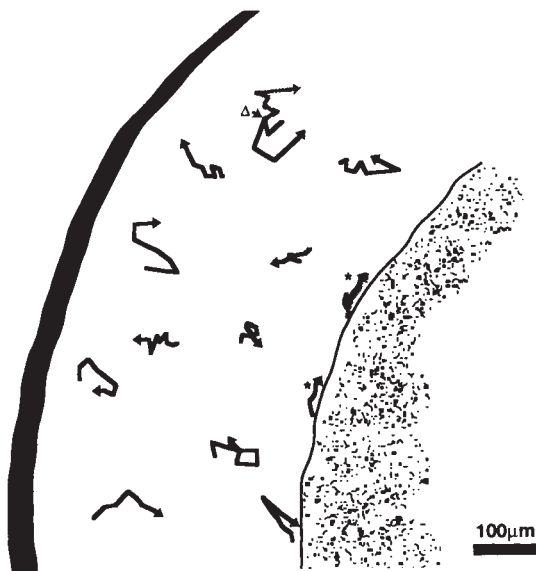


FIG. 4 Path of dispersion of ventricular cells observed over an eight-hour period. The position of Dil-labelled cells in the VZ was recorded by time-lapse video microscopy as described for Fig. 3. A computer reconstruction of the path of dispersion of a sample of these cells, imaged over an 8-h period and compiled from 5 separate preparations, illustrates the movements of a sample of cells. To evaluate whether cells moved in a 'random walk', the mean-square displacement of 21 Dil-labelled cells was plotted as a function of time¹⁴. The linearity of these plots (not shown) indicated that the cells moved in a diffusion mode. The Δ symbol marks the movement of two daughter cells after division of a cell in the plane of focus. The asterisk indicates cells in the border region between the cortical and basal telencephalic germinal zones. In total, cells in 25 explants were imaged, of which 29 Dil-labelled cells were tracked in the border region.

development¹⁸, even crossing functional boundaries¹⁹. Studies on the dispersion of retrovirally labelled progenitor cells^{20,21}, or of dye-labelled cells in living tissue slices²², indicates that although most cells migrate along the radial glial system, some move tangentially within the intermediate zone after leaving the VZ. Together these results suggest that the migratory pathways of immature neural cells are not restricted to the radial plane of the neuroaxis in the early stages of cortical histogenesis.

To determine whether cells move between cortical and basal telencephalic VZs, we imaged Dil-labelled cells along the interface between these regions. When labelled cortical precursor cells entered the boundary region, they moved either in a rostral or caudal direction (Fig. 4), indicating that precursor cells do not cross the border between these two telencephalic VZs. The restriction of cell movement between the dorsal and basal telencephalon seen here is consistent with results showing restricted patterns of gene expression in developing forebrain. Vertebrate homologues of several classes of *Drosophila* genes, including the homeotic gene *forked-head* (*BF-1*; ref. 23), *distal less* (*Dlx* (ref. 24), *Tes-1* (ref. 25)) *achaete scute* (*MASH*²⁵) and *empty spiracles* (*EMX-1*; ref. 27) show patterns of expression that are restricted to either the developing cortex or lateral ganglionic eminence. The segregation of precursor cells into two large domains, the cortical VZ and basal telencephalic VZ, provides a mechanism for the subdivision of the forebrain into two major regions. Although these results superficially resemble the development of the hindbrain, where segmental units (the rhombomeres²⁸) show restricted patterns of lineage, cell patterning and gene expression, there are several distinctions between forebrain and hindbrain development. The zones defined here are much larger than the developing rhombomeres. In addition, the dorsal aspect of the forebrain does not show the segmental patterning or expression of the segment-specific homeotic genes seen in the developing hindbrain²⁹.

The dispersion of neural progenitors within germinal zones of forebrain, observed by imaging dye-labelled cells, supports the general conclusion that cell movements are not rigidly restricted during cortical histogenesis. As the surface area of the VZ is quite small compared with that of the overlying cortical plate, the lateral movement of progenitor cells within the VZ provides a simple mechanism for widespread dispersion of neural progenitor cells in the developing forebrain. Identification of the molecular and cellular signals that establish the borders between cortical and basal telencephalic germinal zones should provide further insight into the mechanisms of pattern formation in the forebrain. □

Received 14 December 1992; accepted 4 February 1993.

1. Ramon Cajal, S. *Histologie du Système Nerveux de l'Homme et des Vertébrés* Vol. II (Maloine, Paris, 1911).
2. Sauer, F. C. J. *comp. Neurol.* **63**, 13–23 (1935).
3. Sauer, M. E. & Chittenden, A. C. *Proc. Soc. exp. Biol. Med.* **101**, 557–560 (1959).
4. Rakic, P. J. *comp. Neurol.* **141**, 283–312 (1971).
5. Sidman, R. L. & Rakic, P. *Brain Res.* **62**, 1–35 (1973).
6. Berry, M. & Rogers, A. W. J. *Anat.* **99**, 691–709 (1965).
7. The Boulder Committee *Anat. Rec.* **166**, 257–262 (1970).
8. Smart, I. H. M. & Sturrock, R. R. *Ontogeny of the Neostriatum* **23**, 127–146 (1979).
9. Honig, M. C. & Hume, R. I. *Trends Neurosci.* **12**, 333–341 (1989).
10. Friederikson, K. & McKay, R. J. *Neurosci.* **58**, 815–821 (1988).
11. Mission, J.-P., Takahashi, R. & Caviness, V. S. *Glia* **4**, 138–148 (1991).
12. Edmondson, J. C. & Hatten, M. E. *J. Neurosci.* **7**, 1928–1934 (1987).
13. Godefont, P., Salaun, J. & Mason, C. A. *Neuron* **5**, 173–186 (1990).
14. Gruler, H. & Bultmann, B. D. *Blood Cells* **10**, 61–77 (1984).
15. Walsh, C. & Cepko, C. L. *Soc. Neurosci. Abstr.* **18**, 925 (1992).
16. Luskin, M. B., Pearlman, A. L. & Sanes, J. R. *Neuron* **1**, 635–647 (1988).
17. Price, J. & Thurlow, L. *Development* **104**, 473–482 (1988).
18. Walsh, C. & Cepko, C. L. *Science* **241**, 1342–1345 (1988).
19. Walsh, C. & Cepko, C. L. *Science* **255**, 434–440 (1992).
20. Misson, J. P., Austin, C. P., Takahashi, T., Cepko, C. L. & Caviness, V. S. *Cerebral Cortex* **1**, 221–234 (1991).
21. Gray, G. E. & Sanes, J. R. *Neuron* **6**, 211–225 (1992).
22. O'Rourke, N. A., Dailey, M. E., Smith, S. J. & McConnell, S. K. *Science* **258**, 299–302 (1992).
23. Porteus, M. H., Bulfone, A., Ciarnello, R. D. & Rubenstein, J. L. R. *Neuron* **7**, 221–229 (1991).
24. Tao, W. & Lai, E. *Neuron* **8**, 957–966 (1992).
25. Price, M., Lemaistre, M., Pischetola, M., Di Lauro, R. & Duboule, D. *Nature* **351**, 748–751 (1991).
26. Lo, L.-C. et al. *Genes Dev.* **5**, 1524–1537 (1991).
27. Simeone, A. et al. *Nature* **358**, 687–690 (1992).
28. Fraser, S., Keynes, R. & Lumsden, A. *Nature* **344**, 431–435 (1990).
29. McGinnis, W. & Krumlauf, R. *Cell* **68**, 283–302 (1992).

ACKNOWLEDGEMENTS. We thank T. M. Jessell, M. Placzek and J. Dodd for comments and critical reading of the manuscript, and C. Walsh, C. L. Cepko and S. K. McConnell for communication of results before publication. Supported by an NIH grant (M.E.H.), an NSF FAW award (M.E.H.) and an MRC of Canada scholarship (G.F.).

Radial mosaicism and tangential cell dispersion both contribute to mouse neocortical development

Seong-Seng Tan & Sibilah Breen

Embryology Laboratory, Department of Anatomy & Cell Biology, The University of Melbourne, Parkville 3052, Australia

THE mammalian neocortex is generated by waves of migrating cells originating from the ventricular zone^{1,2}. Radial migration³ along radial glia^{4,5} has been proposed as the dominant mechanism for this process. The radial unit hypothesis³ is poorly supported by retroviral lineage studies, however, and although some clones show limited radial organization⁶, the emphasis appears to be on widespread tangential dispersion^{7–10}. Here we investigate the pattern of cortical cell dispersion using transgenic mice in which roughly half of the brain cells are coloured by a transgene¹¹. We find that the neocortex is randomly divided into diffused bands, and the majority of cells within each band have the same colour, and their radial orientation suggests radial dispersion. Superimposed upon this was a significant contribution by tangentially dispersed cells that did not respect clonal borders. These observations indicate that cortical specification is not dependent upon a single

mechanism of cell allocation, but that both radial mosaicism and tangential cell migration are involved.

We used transgenic mice for marking progenitor cells with *lacZ*-encoded β -galactosidase, and a genetic mechanism (X-chromosome inactivation) for switching off the gene in half of these cells, generating a functional genetic mosaicism¹¹ (Fig. 1). X inactivation is fully completed by day 12.5 of gestation^{11,12}; blue and white cells in the cerebral wall were therefore considered as stably marked cells. Cell dispersion at this stage was predominantly radial, presenting as alternating blue and white stripes from the ventricular layer towards the primordial plexiform zone (Fig. 2a). In 100- μ m sections, blue and white cells formed radial columns that were occasionally of single-cell diameter but frequently of multiple cell widths. These results suggest that most descendants of progenitor cells have stayed together as a clone; unrestricted mixing at this stage would otherwise have produced random arrangements of blue and white columns of single-cell width.

In adult dorsomedial neocortex, radial bands dominated the landscape but these were superimposed by an even distribution of tangentially dispersed cells. Radial bands have their long axes perpendicular to the brain surface, with ventricular and pial surfaces in register (Fig. 2b). Blue-coloured cells were strongly represented within a blue band along the entire radial length, with the exception of layer I, which had few blue cells, and deep layer VI (derived from subplate), which appeared to be exempt from the banding pattern. Bands had variable rostro-caudal widths (100–1,000 μ m), although it was unclear whether a thick band had arisen from juxtaposition of several thinner bands, or from greater clonal expansion. Measurements of band widths did not indicate quantal differences between thin and thick bands, but quantal increases may have been obscured by non-uniform rates of cell proliferation. Bands appeared ran-

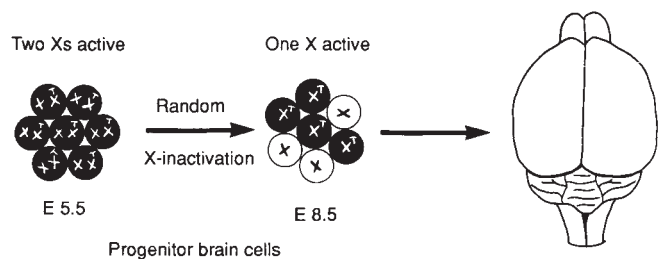
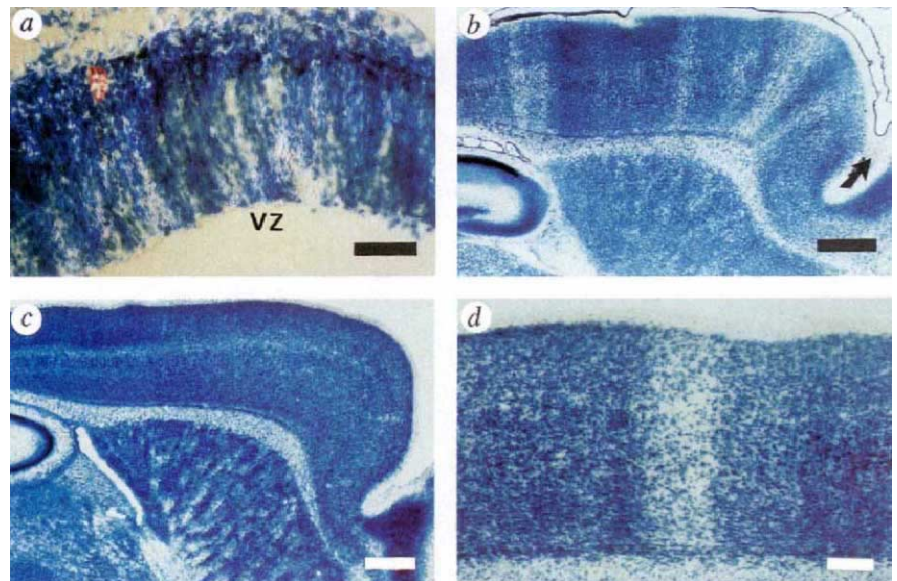


FIG. 1 Method for visualizing mosaicism in the brain following random inactivation of an X-linked *lacZ* transgene. Mosaicism was achieved by the insertion of the transgene into one of the two X chromosomes of line H253 (ref. 11). By embryonic day (E) 8.5, the somatic cell pool from which the brain is derived had randomly inactivated one of the two X chromosomes in female embryos, resulting in a balanced assortment of blue (transgene-active) and white (transgene-inactive) cells. An undetermined number of primordial precursor cells will be withdrawn from this pool to form the embryonic forebrain. Regardless of the number, these precursors will ordinarily make up an equal proportion of expressing and non-expressing cells. Further proliferation and expansion of these precursor cells will give rise to clones that retain their expression phenotype. In the absence of extensive cell mixing, the clonal boundaries of *lacZ*-expressing cells can be revealed by X-gal histochemistry. X^T denotes the transgene on the active X chromosome.

domly placed from rhinal fissure to occipital region; this unpredictability in band size and position was seen from animal to animal and even between right and left halves of individual brains, discounting any correspondence between banding patterns and cytoarchitecture. Instead, their positions are probably the result of random jostling of founder cells during laying down of the embryonic neuroepithelium.

FIG. 2 Migration patterns of β -galactosidase-expressing cell populations in embryonic cerebral wall and adult dorsomedial neocortex of transgenic mice. **a**, Cerebral wall of a E12.5 hemizygous embryo cut in a coronal plane. Parallel blue and white bands of multiple cell thicknesses run in radial alignment from the ventricular zone (VZ) to the earlier formed primordial plexiform zone. Blue cells formed chains of radially aligned cell nuclei. The presence of the blue marker in the nucleus is due to the concentration of β -galactosidase in the cell nucleus¹¹. It should be noted that there appeared to be many more blue cells than white because overlapping white cells tend to be invisible in these preparations. **b**, Alternating blue and white bands delineate clonal boundaries of radially dispersed cells spanning five neocortical layers in adult hemizygous females. These parasagittal sections have been obtained from the dorsomedial neocortex. The arrow is in the rhinal fissure, pointing towards the rostral aspect of the brain. **c**, The control neocortices of homozygous female mice (XX) and male mice (XY) show ubiquitous transgene expression due to an active transgenic allele in every cell. **d**, Blue cells in a white band appear homogeneously distributed in all neocortical layers except for layer 1 and deep layer 6. Scale bar represents 100 μ m in **a**; 500 μ m in **b** and **c**; 200 μ m in **d**.

METHODS. X-gal histochemistry: embryos were obtained by caesarean delivery at 12.5 days post-coitum (E12.5) and their brain tissues were immersion-fixed. Adult brains were obtained after cardiac perfusion, and sectioned on a freezing microtome at 100- μ m thickness. β -galactosidase activity in embryonic and adult brain sections was assayed as described¹¹. A total of 49 hemizygous females, 4–8 weeks old, were used for the study. Two litters provided 8 transgenic embryos at E12.5. Control brains came



from 10 males and 2 homozygous females. Numerical estimation of blue and white cells: after X-gal histochemistry of parasagittal sections, blue ($n=14$) and white ($n=15$) bands were dissected from parasagittal sections of two brains and embedded in paraffin for thin sectioning (10 μ m) in their long axes. Every third section in a band was sampled, and the percentage of blue cell nuclei in a blue band (and corresponding white cells in a white band) was calculated from a total of 188 sections (101 blue, 87 white) after counterstaining in nuclear fast red. Comparisons were made between blue and white areas from within an animal, and between two animals using the non-parametric Mann-Whitney test.

To eliminate the possibility of bands having resulted from capricious transgene expression, control experiments were performed with homozygous females. These have the transgene on both X chromosomes and their neocortices showed uniform expression (Fig. 2c). Ubiquitous expression was also seen in the cortices of male transgenic animals with a single X chromosome, confirming the validity of using X inactivation for generating phenotypic mosaicism.

Radial bands were, however, not homogeneously only blue or white. This was more obvious in white areas where tangentially dispersed blue cells were evenly distributed in five cortical layers (Fig. 2d). Our present data do not permit discussion on the identities and fates of tangentially dispersing cells, but it would seem that both neurons and glia are involved (manuscript in preparation). To ascertain the degree of tangential dispersion, individual blue and white bands were dissected out from parasagittal sections of two animals, re-embedded in paraffin, and the proportion of blue cells situated within a blue band (and corresponding white cells in a white band) counted. The results from both types of band were uniform and complementary, showing that about two-thirds of cells in a band were the same colour (that is, percentage of the blue cells in blue bands (mean $\pm$ s.e.m.) was 64.9 ± 1.8 , $n = 14$ bands; white cells in white bands, 67.3 ± 1.67 , $n = 15$ bands; non-parametric Mann-Whitney test $P < 0.05$). To see whether cells of the other colour were evenly distributed in a band, we used the optical dissector method¹³ to compare the colour distribution profiles in middle and side areas of white bands. We found that the middle and side areas of 500 μ m bands had similar blue/white ratios of cells. At least two conclusions can be drawn from these experiments. One is that most of the cells in a band have dispersed radially (as seen in the ferret¹⁴), with the remaining one-third of cells having the other colour, representing horizontal migration of progenitor cells in the ventricular zone, causing overlap between adjacent clones. Future comparisons between thick and thin bands may throw more light on this issue. Alternatively, only one third of cells in this study have remained stationary (and radially dispersing), whereas two-thirds (one-third blue and one-third white) have entered a band by tangential dispersion. This assumes that tangentially dispersing cells have unlimited capacities for cell dispersion and are able to cross multiple clonal boundaries¹⁰.

In conclusion, the neocortex is not constructed by a single mechanism of cell dispersion. Both radial mosaicism of clonally restricted cells and tangential dispersion of freely mixing cells are involved and may represent separate facets of cortical development. Tangential dispersion of some cells would endow the developing neocortex with a high degree of plasticity^{15,16} by postponing commitment until arrival at their final locations. Radially deployed cells, on the other hand, have their formal positions fixed by their starting locations in the ventricular layer, therefore their final positions in the neocortex depend on the original neuroepithelial mosaicism. Radial mosaicism means that the neuroepithelium has subpopulations of cells that, amid widespread tangential dispersion, are capable of generating diffused radial columns in register with their place of origin. This early organizational feature may be important in providing scaffolding upon which the growing neocortex can be built; this is in addition to other autonomous properties of the neuroepithelium, such as the intrinsic ability to generate a laminated cortex, and possession of a common circuitry for information processing^{17,18}. □

Received 3 September 1992; accepted 11 February 1993.

1. Angevine, J. B. & Sidman, R. L. *Nature* **192**, 766–768 (1961).
2. Rakic, P. *Science* **183**, 425–427 (1974).
3. Rakic, P. *Science* **241**, 170–176 (1988).
4. Rakic, P. *J. comp. Neurol.* **141**, 283–312 (1971).
5. Gadisseuz, J. F., Evrad, P., Misson, J. P. & Caviness, V. S. Jr *Devl Brain Res.* **50**, 55–67 (1989).
6. Luskin, M. B., Pearlman, A. L. & Sanes, J. R. *Neuron* **1**, 635–647 (1988).
7. Walsh, C. & Cepko, C. L. *Science* **241**, 1342–1345 (1988).

8. Austin, C. P. & Cepko, C. L. *Development* **110**, 713–732 (1990).
9. Price, J. & Thurlow, L. *Development* **104**, 473–482 (1988).
10. Walsh, C. & Cepko, C. L. *Science* **255**, 434–440 (1992).
11. Tan, S.-S., Williams, E. A. & Tam, P. P. L. *Nature Genet.* **3**, 170–174 (1993).
12. Monk, M. & Harper, M. I. *Nature* **281**, 311–313 (1979).
13. Gundersen, H. J. G. et al. *Acta path. microbiol. immunol. scand.* **96**, 857–881 (1988).
14. O'Rourke, N. A., Dailey, M. E., Smith, S. J. & McConnell, S. K. *Science* **258**, 299–302 (1992).
15. Molnár, Z. & Blakemore, C. *Nature* **351**, 475–477 (1991).
16. Schlaggar, B. L. & O'Leary, D. D. *M. Science* **252**, 1556–1560 (1991).
17. Sur, M., Garraghty, P. E. & Roe, A. W. *Science* **242**, 1437–1441 (1988).
18. Métin, C. & Frost, D. O. *Proc. natn. Acad. Sci. U.S.A.* **86**, 357–361 (1989).

ACKNOWLEDGEMENTS. We thank C. Cepko, R. Guillery, J. Price, P. Rakic and C. Walsh for criticism; J. Furness, C. Masters and I. Darian-Smith for discussion; M. Mehtali for plasmid p11-HMGpro; and F. Weissenborn and D. Wong for technical assistance. This work was supported by the NH&MRC, and the George Hicks, William Buckland and Ian Potter Foundations.

Phosphorylation and regulation of glutamate receptors by calcium/calmodulin-dependent protein kinase II

Ellen McGlade-McCulloh, Hideyuki Yamamoto, Soon-Eng Tan, Debra A. Brickey & Thomas R. Soderling*

Vollum Institute, Oregon Health Sciences University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97201, USA

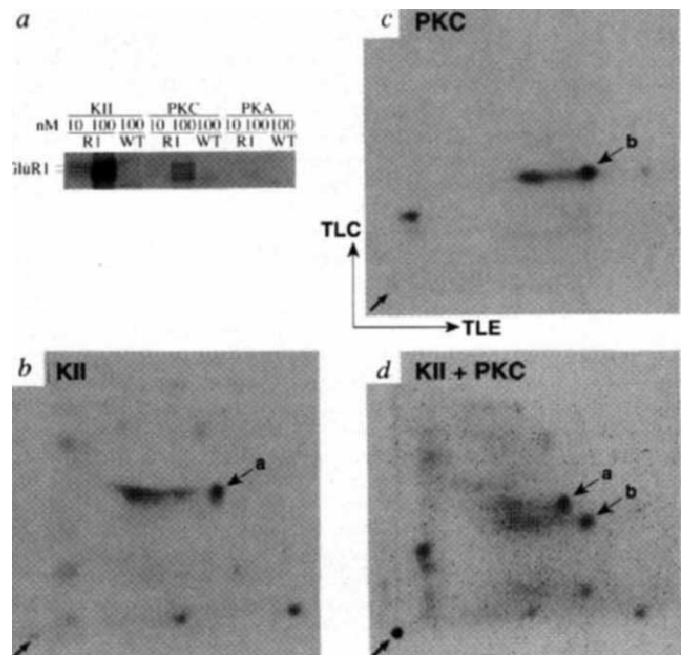
THE major postsynaptic density (PSD) protein^{1,2} at glutaminergic synapses is calcium/calmodulin-dependent protein kinase II (CaM-K II), but its function in the PSD is not known. We have examined glutamate receptors (GluRs) as substrates for CaM-K II because (1) they are colocalized in the PSD³, (2) cloned GluRs^{4–7} contain consensus phosphorylation sites for protein kinases including CaM-K II, and (3) several GluRs are regulated by other protein kinases^{8–12}. Regulation of GluRs, which are involved in excitatory synaptic transmission and in mechanisms of learning and memory¹³, by CaM-K II is of interest because of the postulated role of CaM-K II in synaptic plasticity^{14,15} and its known involvement in induction of long-term potentiation¹⁶. Furthermore, mice lacking the major neural isoform of CaM-K II exhibit deficits in models of learning and memory that require hippocampal input^{17,18}. We report here that CaM-K II phosphorylates GluR in several *in vitro* systems, including the PSD, and that activated CaM-K II enhances kainate-induced ion current three- to fourfold in cultured hippocampal neurons. These results are consistent with a role for PSD CaM-K II in strengthening postsynaptic GluR responses in synaptic plasticity.

Initial experiments examined the phosphorylation of homomeric GluR1, expressed using the baculovirus/Sf9 insect cell system, which was purified by immunoprecipitation with an antibody specific for its carboxy terminus. On addition of 100 nM purified protein kinases, GluR1 was rapidly phosphorylated by CaM-K II, slowly phosphorylated by protein kinase C (PKC) and not detectably phosphorylated by protein kinase A (PKA) (Fig. 1a). These results are consistent with the presence of consensus phosphorylation sites in the intracellular loop of GluR1–4 for CaM-K II and PKC but not for PKA^{4,5}. Phosphoaminoacid analyses (data not shown) established that CaM-K II and PKC predominantly phosphorylated Ser with only 13% and 27%, respectively, of the radioactivity in Thr. Two-dimensional ³²P-peptide maps for GluR1 phosphorylated by CaM-K II or PKC are shown in Fig. 1b–d. The major CaM-K II phosphorylation site (Fig. 1b, peptide a) was different from the major PKC phosphorylation site (Fig. 1c, peptide b) as clearly shown

* To whom correspondence should be addressed.

FIG. 1 Phosphorylation of GluR1 expressed in Sf9 cells. *a*, Immunoprecipitates, using an antibody (Ab) to GluR1, from control (WT) or GluR1 (R1)-expressing Sf9 cells were 32 P-labelled using 10 or 100 nM of purified CaM-K II (KII), PKC, or PKA and subjected to SDS-PAGE and autoradiography. The positions of the nonglycosylated 91K and glycosylated 102K GluR1 are indicated to the left. *b-d*, Two-dimensional tryptic 32 P-peptide maps of Sf9 GluR1 phosphorylated for 30 min by 100 nM CaM-K II (*b*), PKC (*c*) or both (*d*). The arrows *a* and *b* designate the major phosphopeptides; arrows in lower left corner of each panel indicate the origins.

METHODS. Cultured Sf9 cells²⁵ were infected for 3 days with wild-type or recombinant baculovirus for GluR1 (complementary DNA provided by Hollmann and Heinemann, Salk Institute). The 10,000g supernatant fractions from wild-type (WT) or GluR1-expressing (R1) Sf9 cells were incubated with GluR1 Ab²³ (provided by R. Wenthold, NIH) overnight at 4 °C followed by incubation with Protein-A Sepharose at 4 °C for 2 h. After centrifugation at 10,000g, the pellets were washed four times with 0.01% BSA, 100 mM Tris (pH 7.5), 0.9% NaCl. Immunoprecipitates were phosphorylated with 10 or 100 nM purified CaM-K II α subunit (KII)²⁷, PKC (provided by K. P. Huang, NIH), or PKA catalytic subunit (provided by J. D. Corbin, Vanderbilt) for 3 min at 30 °C, reprecipitated, washed and run on SDS-PAGE. The phosphorylation reactions contained 50 mM HEPES (pH 7.5), 10 mM magnesium acetate, 0.1 mM [γ - 32 P]ATP plus the following special conditions for each indicated kinase: CaM-K II, 1 mM CaCl₂, 6 μ M CaM; PKC, 1 mM CaCl₂, 50 μ g ml⁻¹ phosphatidylserine, 5 μ g ml⁻¹ 1,3-diolein; PKA, 1 mM EGTA and 50 mM potassium phosphate buffer (pH 6.8) in place of HEPES. *b-d*, The phosphorylated (30 min, 100 nM K II or PKC) immune complex was recentrifuged and subjected to SDS-PAGE. The 102K 32 P-GluR1 was excised from the gel and digested with trypsin (100 μ g TPKC-trypsin in 50 mM NH₄HCO₃, pH 8, overnight at 37 °C, then another 100 μ g for 3 h). After centrifugation, the supernatants were lyophilized and subjected to two-dimensional peptide mapping using the Hunter Thin Layer Electrophoresis system (HTLE-7000, CBS Scientific Co). *b*, GluR1 phosphorylated by CaM-K II; *c*, GluR1 phosphorylated by PKC; *d*, mixture of GluR1 phosphorylated by CaM-K II or PKC.



when the two samples were mixed before separation (Fig. 1d).

Native GluR was also strongly phosphorylated by endogenous CaM-K II in preparations of nerve terminals and PSDs. The PSD, a specialized organelle tightly apposed to the postsynaptic membrane of forebrain glutamatergic synapses, is thought to be involved in a long-term regulation of synaptic responses¹⁹. Synaptosomes contain endogenous CaM-K II, PKC and PKA (data not shown) which were selectively activated to ascertain their 32 P-labelling of GluRs as detected by subsequent immunoprecipitation with GluR1 antibody (Fig. 2, top). Compared to basal phosphorylation of GluR (defined as 100%) in the presence of EGTA (lane 4), CaM-K II (lane 2) strongly enhanced (474%) phosphorylation of GluR, PKC (lane 6) gave a weaker (168%) response, whereas PKA (lane 8) resulted in little if any (114%) enhanced 32 P-labelling of GluR. When isolated PSDs were used as the source of GluRs and endogenous

protein kinases (lanes 2, 4, 6), only CaM-K II catalysed strong phosphorylation (Fig. 2, bottom, lane 2); this was not surprising because CaM-K II constitutes 30% of the PSD protein^{1,2}. If we added 100 nM of each purified protein kinase to the PSD (lanes 1, 3, 5), the exogenous PKC (lane 3) slightly enhanced phosphorylation, whereas purified PKA (lane 5) did not catalyse significant phosphorylation. The lack of PKA phosphorylation of GluR in all *in vitro* systems examined suggests that the effect of PKA on kainate/2-(aminomethyl)phenylacetic acid (AMPA) current observed in cultured hippocampal neurons^{9,10} may be due to the phosphorylation of GluR²⁰ which does contain a consensus PKA phosphorylation site. The slow phosphorylation of GluR by PKC may have physiological relevance because the proteolytically activated form of PKC can enhance hippocampal kainate/AMPA current¹².

Because CaM-K II strongly catalysed *in vitro* phosphorylation of GluR, we investigated whether CaM-K II can regulate GluR

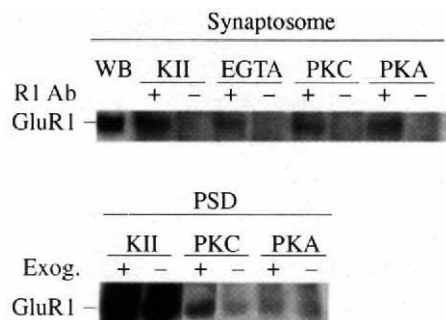


FIG. 2 Autoradiographs of synaptosomal and PSD GluR1 phosphorylated by protein kinases. Top, Isolated rat brain synaptosomes²⁸ were lysed by hypotonic freezing-thawing and phosphorylated for 3 min at 30 °C using endogenous protein kinases in the presence of selective kinase activators/inhibitors or EGTA. After solubilization, immunoprecipitates were

made using antibody to GluR1 (R1 Ab+) or control serum (R1 Ab-), and subjected to SDS-PAGE/autoradiography. The left lane is a western blot (WB) using GluR1 Ab and ¹²⁵I-Protein A. Bottom, Isolated rat forebrain PSDs²⁸ were phosphorylated using selective kinase activation conditions and endogenous protein kinases (Exog. -) or with 100 nM exogenous purified kinases (Exog. +) as indicated. PSDs were solubilized, precipitated with GluR1 Ab, and subjected to SDS-PAGE/autoradiography.

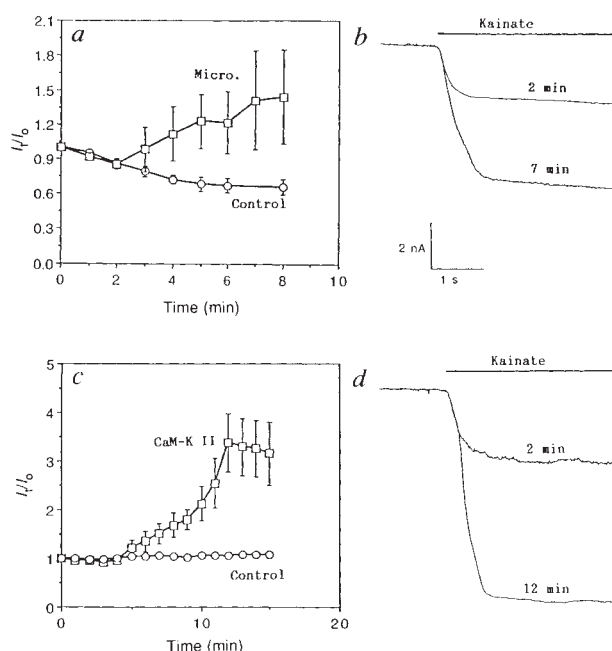
METHODS. Phosphorylation reactions contained 50 mM HEPES (pH 7.5), 10 mM magnesium acetate, 0.4 mM [γ - 32 P]ATP plus the following selective kinase activators/inhibitors. CaM-K II: 0.5 mM CaCl₂, 2 μ M CaM, 5 μ M PKA-inhibitor peptide (PKA-IP), 2 μ M PKC-IP; EGTA control: 1 mM EGTA, 5 μ M PKA-IP, 2 μ M PKC-IP; PKC: 0.5 mM CaCl₂, 3 μ M mastoparan (to inhibit CaM-K II), 50 μ g ml⁻¹ phosphatidylserine, 5 μ g ml⁻¹ 1,3-diolein, 5 μ M PKA-IP; PKA: 10 μ M cAMP, 1 mM EGTA, 3 μ M mastoparan, 2 μ M PKC-IP. The synaptosomes or PSDs were solubilized with SDS buffer containing 1% SDS, diluted 10-fold with 50 mM Tris, 150 mM NaCl, 10 mM EDTA, 2 mM EGTA, 50 mM NaF, 0.5 μ M microcystin-LR, 1% Nonidet P-40, 20 μ g ml⁻¹ soybean trypsin inhibitor, 5 μ g ml⁻¹ leupeptin, and 1 mM phenylmethylsulphonyl fluoride (PMSF), pH 7.5, and centrifuged at 10,000g. Supernatants were immunoprecipitated overnight with GluR1 Ab as in Fig. 1. The specificities of the PKA-IP and PKC-IP have been reported²⁹.

FIG. 3 Effects of a protein phosphatase inhibitor (*a*, micro and *b*) or activated CaM-K II (*c*, *d*) on hippocampal whole-cell, kainate-induced inward current. *a*, *c*, Hippocampal neurons were perfused for 2–4 s with 100 μ M kainate at the indicated times in the presence of 10 μ M microcystin-LR (*a*, micro.) or 200 nM activated CaM-K II (*c*) in the pipette solution. The ratio of the maximum kainate-induced current at the indicated time (I_t) was divided by the maximum kainate-induced current attained shortly after patch formation (I_0) at a holding potential of -60 mV. Data are presented as mean $\pm$ s.e.m. ($n=3-4$ for micro.; $n=4-7$ for CaM-K II). *b*, *d*, Representative kainate-induced currents at the indicated times with 10 μ M microcystin (*b*) or 200 nM activated CaM-K II (*d*) in the pipette. Neither microcystin or CaM-K II altered the leakage current.

METHODS. Rat hippocampal neurons were isolated and cultured as described³⁰. External solution: 160 mM NaCl, 5 mM KCl, 2 mM CaCl_2 , 10 mM HEPES (pH 7.4), 10 mM glucose, 0.5 μ M tetrodotoxin. Pipette solution for *a* and *b*: 140 mM caesium methylsulphonate, 5 mM CsCl, 11 mM EGTA, 1 mM CaCl_2 , 5 mM Mg^{2+} -ATP, 5 mM HEPES (pH 7.4). Pipette solution for *c* and *d*: 140 mM sodium gluconate, 5 mM NaCl, 11 mM EGTA, 1 mM CaCl_2 , 5 mM Mg^{2+} -ATP, 5 mM HEPES (pH 7.4). Kainate was perfused onto the neuron by manual changing of a double-barrelled pipette positioned 50 μ m from the cell. CaM-K II, α subunit expressed in Sf9 cells and purified²⁸, was stably activated (60–70% Ca^{2+} -independence) by autophosphorylation. CaM-K II (2 μ M) was incubated for 10 min at 5 $^\circ\text{C}$ in 50 mM HEPES (pH 7.5), 0.5 mM CaCl_2 , 6 μ M CaM, 10 mM magnesium acetate, 0.4 mM ATP- γ -S, and 1 mg ml^{-1} BSA. The autophosphorylated CaM-K II was maintained on ice and diluted 10-fold in the pipette solution just before use. Because of the thermal instability of the autophosphorylated CaM-K II, kinase activity of parallel diluted samples were monitored. For the control in *c*, the CaM-K II was heat-inactivated (10 min at 100 $^\circ\text{C}$) before addition to the autophosphorylation reaction.

ion current in cultured hippocampal neurons. We first examined the effect of protein phosphatase inhibitors on kainate current because the activation state of CaM-K II can be increased in neurons by a phosphatase inhibitor²¹. Standard whole-cell patch clamp recording methods were used to study kainate responses in 6–14 day cultured hippocampal neurons. Kainate, rather than glutamate, was chosen as agonist as it does not exhibit desensitization. The 'run down' of kainate-induced current with repeated agonist application was largely prevented by inclusion of Mg^{2+} /ATP in the pipette^{9,10} (Fig. 3*a*, control). Addition of the protein phosphatase inhibitor microcystin-LR²² (10 μ M), enhanced the kainate current 1.5–2-fold in a time-dependent manner (Fig. 3*a* micro and *b*). More importantly, intracellular application of CaM-K II (200 nM), activated by autophosphorylation to give 60–70% Ca^{2+} -independent activity, enhanced kainate-induced current three- to fourfold compared to the heat-inactivated CaM-kinase II control (Fig. 3*c*, *d*). ATP- γ -S was used in the autophosphorylation because thiophosphorylated proteins are more resistant to protein phosphatases, which can rapidly dephosphorylate soluble CaM-kinase II in neurons²¹. In two experiments, inclusion of nonphosphorylated CaM-K II (200 nM) in the pipette gave no significant effect on kainate-induced current above Mg^{2+} /ATP alone (data not shown). This is not surprising as the CaM-K II would not be expected to be active in the absence of added calmodulin and at the low intracellular Ca^{2+} used in these experiments. These results with CaM-K II and protein phosphatase inhibitors indicate that the AMPA/kainate ion channels are strongly modulated by their phosphorylation state.

We have demonstrated here that CaM-K II catalysed the *in vitro* ^{32}P -labelling of GluR and enhanced kainate current in hippocampal neurons. Because the CaM-K II consensus phosphorylation site in GluR1 (Ser627) is conserved in all AMPA/kainate GluRs cloned so far, it is likely that other members of this family are also phosphorylated by CaM-K II and may contribute to the enhanced kainate current response. Our results provide a rationale for the colocalization of CaM-K II and GluRs in the PSD of forebrain excitatory synapses, and they suggest that CaM-K II enhances kainate current by direct phosphorylation of GluRs rather than of some regulatory protein. One hypothesis that results from this work is that critical



synaptic events that stimulate postsynaptic Ca^{2+} influx, such as NMDA (*N*-methyl-D-aspartate) receptor stimulation, produce activation of CaM-K II through its autophosphorylation. This Ca^{2+} -independent CaM-K II in the PSD could then phosphorylate AMPA/kainate GluRs resulting in enhanced postsynaptic currents in response to subsequently released glutamate. In support of this hypothesis, we have recently demonstrated that NMDA receptor activation in cultured hippocampal neurons does generate the Ca^{2+} -independent, activated CaM-K II²⁶. Thus, our experiments are consistent with a role for CaM-K II as a molecular switch or sensor of synaptic activity as previously postulated^{14,15}. Further work will be required to determine if phosphorylation of GluRs by activated CaM-K II contributes to the enhanced postsynaptic AMPA/kainate response in long-term potentiation²⁴ or other models of learning and memory. □

Received 26 August 1992; accepted 1 February 1993.

- Kennedy, M. B., Bennett, M. K. & Erndu, N. G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7357–7361 (1983).
- Kelly, P. T., McGuinness, T. L. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 945–949 (1984).
- Wu, K., Carlin, R. & Siekevitz, P. *J. Neurochemistry* **46**, 831–841 (1986).
- Boulter, J. *et al. Science* **249**, 1033–1037 (1990).
- Keinänen, K. *et al. Science* **249**, 556–560 (1990).
- Egebjerg, J., Bettler, B., Hermans-Borgmeyer, I. & Heinemann, S. *Nature* **351**, 745–748 (1991).
- Moriyoshi, K. *et al. Nature* **354**, 31–37 (1991).
- Knapp, A. G., Schmidt, K. F. & Dowling, J. E. *Proc. natn. Acad. Sci. U.S.A.* **87**, 767–771 (1990).
- Wang, L.-Y., Salter, M. W. & MacDonald, J. F. *Science* **253**, 1132–1135 (1991).
- Greengard, P., Jen, J., Nairn, A. C. & Stevens, C. F. *Science* **253**, 1135–1138 (1991).
- Keller, B. U., Hollmann, M., Heinemann, S. & Konnerth, A. *EMBO J.* **11**, 891–896 (1992).
- Wang, L. Y., Dudek, E. M., Browning, M. & MacDonald, J. F. *Neurosci. Abstr.* **18**, 653 (1992).
- Westbrook, G. L. & Jahr, C. E. *Semin. Neurosci.* **1**, 103–114 (1989).
- Lisman, J. E. & Goldring, M. A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5320–5324 (1988).
- Miller, S. G. & Kennedy, M. B. *Cell* **44**, 861–870 (1986).
- Malinow, R., Schulman, H. & Tsien, R. W. *Science* **245**, 862–866 (1989).
- Silva, A. J. *et al. Science* **257**, 201–206 (1992).
- Silva, A. J. *et al. Science* **257**, 206–210 (1992).
- Siekevitz, P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3494–3498 (1985).
- Raymond, L. A., Blackstone, C. D. & Huganir, R. L. *Nature* **361**, 637–641 (1993).
- Fukunaga, K. & Soderling, T. R. *Molec. cell. Neurosci.* **1**, 133–138 (1990).
- MacKintosh, C. *et al. FEBS Lett.* **264**, 187–193 (1990).
- Westhold, R. J., Yokotani, N., Doi, K. & Wada, K. *J. Biol. Chem.* **267**, 501–507 (1992).
- Manabe, T., Renner, P. & Nicoll, R. A. *Nature* **355**, 50–55 (1992).
- Summers, M. D. & Smith, G. E. *Texas Agricultural Experiment Station Bulletin No.* 1555 (1988).
- Fukunaga, K., Soderling, T. R. & Miyamoto, E. *J. Biol. Chem.* **267**, 22527–22533 (1992).
- Bricey, D. A., Colbran, R. J., Fong, Y. L. & Soderling, T. R. *Biochem. biophys. Res. Commun.* **173**, 578–584 (1990).
- Rich, D. P., Colbran, R. J., Schworer, C. M. & Soderling, T. R. *J. Neurochem.* **53**, 807–816 (1989).
- Smith, M. K., Colbran, R. J. & Soderling, T. R. *J. Biol. Chem.* **265**, 1837–1840 (1990).
- Lester, R. A. J. *et al. Molec. Pharmacol.* **35**, 565–570 (1989).

ACKNOWLEDGEMENTS. We thank C. Jahr and G. Westbrook for providing cultured hippocampal neurons and colleagues at the Vollum Institute for helpful suggestions on this manuscript. Supported by a grant from the NIH (to T.R.S.).

A nutrient-permeable channel on the intraerythrocytic malaria parasite

Sanjay A. Desai*†, Donald J. Krogstad*†‡
& Edwin W. McCleskey†

Departments of †Cell Biology, ‡Medicine, Pathology and Mechanical Engineering, Washington University, St Louis, 660 S. Euclid Avenue, St Louis, Missouri 63110, USA

DURING its 48-hour cycle inside the red blood cell, the human malaria parasite, *Plasmodium falciparum*, increases its volume 25-fold and divides asexually. This rapid growth demands large amounts of nutrients, a problem exacerbated by the lower metabolic rate and relative ionic impermeability of the host red blood cell. Direct passage of small nutrients across the two membranes that separate the parasite from the erythrocyte cytosol may be important for parasite development¹ and has been demonstrated for radiolabelled glucose², amino acids^{3,4} and purine nucleosides⁴⁻⁶. Flux studies on plasmodia are limited, however, to suspensions of erythrocyte-free parasites and so cannot be used to examine the individual transport properties of the two membranes involved. Here we use the cell-attached patch clamp method⁷ to overcome this limitation. After removing the intervening red blood cell membrane and forming gigaohm seals on the small (3–5 μm) parasite, we studied transport across the parasitophorous vacuole membrane (PVM), the outer of the two membranes that separate the parasite from the erythrocyte cytosol. A 140-pS channel which is permeable to both cations and anions was identified on the PVM. This channel is present at high density, is open more than 98 per cent of the time at the resting potential of the PVM, and is permeable to lysine and glucuronate. The channel can readily transport amino acids and monosaccharides across the PVM and may be essential for fulfilling the parasite's metabolic demands.

The PVM originates as an invagination of the red blood cell (RBC) membrane during invasion and undergoes dramatic changes in composition⁸ with the addition of parasite-synthesized lipids. The specific transport properties of the PVM have been difficult to characterize because it is closely juxtaposed to the underlying parasite plasma membrane and because of the inevitable membrane leakiness of erythrocyte-free parasite suspensions⁹. We used single malaria parasites and the on-cell patch clamp method to record the opening of individual ion channels on the PVM.

Two different methods were used to remove the RBC membrane and expose the intraerythrocytic parasite. The first, demonstrated with sequential photographs (Fig. 1A), used brief exposure to 40 $\mu\text{g ml}^{-1}$ digitonin, a detergent that acts by binding cholesterol¹⁰. Digitonin rapidly permeabilized the cholesterol-rich RBC membrane to produce haemoglobin-free ghosts (two are seen at upper left and one at upper right in *b* of Fig. 1A). When a parasitized RBC is exposed to digitonin, the intraerythrocytic parasite is often extruded from its permeabilized host RBC (parasite seen in mid-extrusion in *c* of Fig. 1A). Electron microscopy demonstrates that the PVM remains intact after this treatment (Fig. 1B): examination of 18 extruded parasites in this manner revealed only intact PVMs without the characteristic membrane damage induced by digitonin in susceptible membranes¹⁰. Subsequent patch clamp of the liberated parasite by sealing a recording pipette on the outer surface of the PVM (Fig. 1A, *d*) revealed unambiguous channel events (Fig. 1C).

The second method of exposing the parasite used brief 200 mV pulses (<10 ms) applied to a patch pipette sealed on the RBC

membrane. This manoeuvre disrupts the RBC membrane irreversibly. Simultaneous suction (~10 cm H₂O) then lodged the parasite inside the pipette's opening, where a gigaohm seal occasionally formed (<5% of parasites). This method revealed the same channel activity as the digitonin method, but produced more durable and higher resistance seals (Fig. 1D). The identical channel activity identified by both methods argues strongly against an artefact induced by either permeabilization with digitonin or the application of electrical pulses. Furthermore, nonspecific permeability changes in the PVM of freed parasites, a major concern with radiolabelled nutrient flux studies¹⁻⁵, could not produce these unambiguous single-channel events.

With Cs⁺ and Cl⁻ as the predominant ions in the pipette solution, both methods revealed a channel with a slope conductance of 140 pS and an apparent reversal potential, E_{rev} , of -2 mV (+2 mV applied to pipette; Fig. 1E). To determine which ions carry current when this channel is open, we replaced the CsCl with glycine, a neutral zwitterion. If Cl⁻ were the only permeant ion in the pipette, the outward anion gradient imposed by reducing the pipette Cl⁻ to 11 mM would shift E_{rev} to +60 mV ($\Delta E_{\text{rev}} = 60 \log(11/11)$). If, on the other hand, only Cs⁺ were permeable, E_{rev} would shift to a similar but negative value because of the imposed outward cation gradient. We found that these outward gradients offset each other: E_{rev} with glycine was +5 mV (Fig. 1E), indicating high permeability to both Cs⁺ and Cl⁻. The small, but significant, shift in E_{rev} from -2 mV to +5 mV might reflect a slightly greater permeability to anions than cations. With glycine, the slope conductance decreased to 85 pS, consistent with a reduction in open channel flux due to lower concentrations of charged substrates.

The permeabilities of other cations and anions were examined by substituting for Cs⁺ or Cl⁻ in the pipette solution: an impermeant ion would shift E_{rev} by at least 20 mV (negative with an impermeant cation, positive with an impermeant anion). Five substitutions permitted gigaohm seals and revealed measurable channel activity (Fig. 2a). These ions (lysine⁺, Tris⁺, Ca²⁺, Mg²⁺ and glucuronate⁻) were all highly permeant because their E_{rev} values were indistinguishable from those of CsCl (Fig. 2b; compare with the solid line, which is the line fitted to CsCl in Fig. 1E). Because lysine⁺ and Tris⁺ are freely permeable in this channel but cannot pass through the 6.5 Å pore formed by the acetylcholine receptor¹¹, this channel's minimum diameter must be greater than 6.5 Å. Moreover, the permeation of large cations and anions suggest similarities to porin channels which have pore diameters of about 11 Å (ref. 12).

In addition to its primary 140 pS conductance (Figs 1E and 2), this channel frequently exhibits subconductance states, openings which allow decreased ion flux. The upper sweep in Fig. 3, recorded at 60 mV below the resting membrane potential, shows transitions to both the primary conductance state (dotted line) and a flickering subconductance state (dashed line). Were this flicker simply noise or an independent channel, it would not disappear with every transition to the primary conductance. This subconductance level, nearly resolved with a faster recording speed in the lower sweep in Fig. 3 (upper dashed line), had a slope conductance of 25 pS and a E_{rev} indistinguishable from that of the primary conductance (not shown). Another less commonly observed subconductance state (lower dashed line, lower sweep) also had the same E_{rev} . The brief transition (arrow) to the fully closed state (solid line) argues that this second subconductance is not an independent outward channel, which would require the unlikely simultaneous closing of two channels¹³. Although similar subconductance behaviour has been seen with other highly permeant channels, its physiological significance is uncertain.

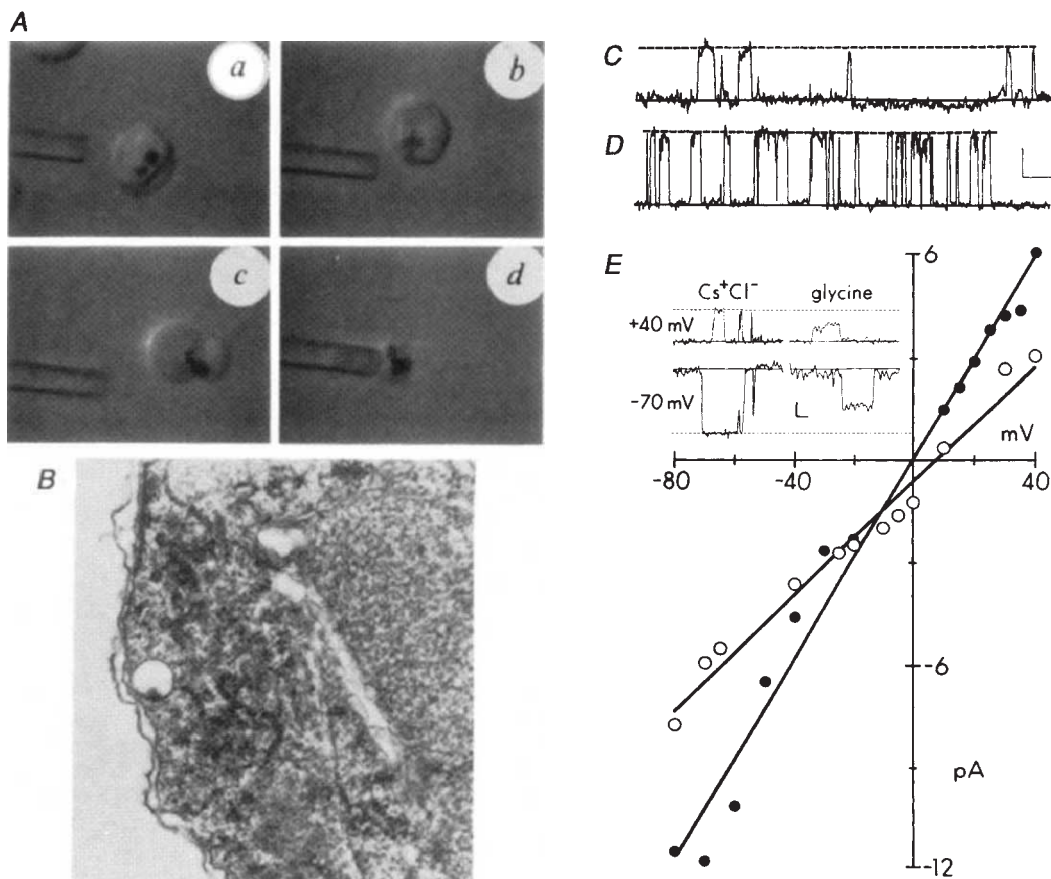
Two observations suggest that this highly permeable channel serves an important physiological role. First, it must have a relatively high density in the PVM as multiple channels were seen on almost all patches. For example, at least seven channels are present in the patch shown in Fig. 4. Second, at the PVM's

* Present addresses: Department of Medicine, Box 31215, Duke University Medical Center, Durham, NC 27710 (S.A.D.); Tulane School of Public Health and Tropical Medicine, 1501 Canal Street, New Orleans, LA 70112, USA (D.J.K.).

FIG. 1 Identification of a 140 pS channel on the PVM.

A, Sequential micrographs demonstrating the digitonin method of patch clamping the parasite: *a*, Parasitized RBC (centre, 7 μ m diameter) and patch pipette (left). Note haemozoin crystals (digested haemoglobin) in parasite; *b*, immediately after a ~ 10 s exposure to 40 μ g ml $^{-1}$ digitonin. The digitonin solution and digitonin-free bath solution for subsequent washout were applied from a distance of ~ 100 μ m using a Drummond 1 μ l micro-pipette connected to solution reservoirs²⁶. Uninfected RBCs become ghosts (two at upper left, one at upper right) within seconds of exposure to digitonin; *c*, the parasite is seen being extruded from a digitonin-permeabilized host cell; *d*, once externalized, the PVM can be patch clamped. Notice that the host RBC remains as a ghost just above the parasite. Patch clamp of host RBC membranes, either before or after digitonin treatment, does not reveal the nutrient channel, confirming that it is unique to the PVM. **B**, Elec-

tron micrograph of an erythrocyte-free parasite prepared by digitonin treatment. Trophozoite-infected RBCs were incubated in bath solution with digitonin for 1 min, centrifuged at 600g for 5 min, and fixed in 1% glutaraldehyde with 0.1 M sodium cacodylate for 15 min at 4 °C. After washing, cells were post-fixed with 1.25% osmium tetroxide for 45 min at room temperature, rinsed and stained in 4% aqueous uranyl acetate for 45 min. Cells were then dehydrated in ethanol and propylene oxide, embedded in Polybed 812, and sectioned for viewing in a Zeiss 10 A electron microscope. Magnification, 80,000 $\times$. **C**, Patch clamp recording obtained with the digitonin method. Dashed and solid lines represent channel's open and closed states, respectively. Pipette solution (mM): 100 CsCl with buffer A (40 HEPES, 10 EGTA, 2.9 CaCl $_2$, 2.5 MgCl $_2$, 1 Na $_4$ ATP, 0.3 GTP, 8.8 phosphocreatine, to pH 7.0 with TEA-OH). Buffer A was designed to simulate RBC cytosol, had free [Ca $^{2+}$] and [Mg $^{2+}$] buffered to 0.1 μ M and 2.0 mM with EGTA, respectively, and was not essential for channel activity. V_p (negative of the voltage applied to the pipette) was +25 mV. Scale is 2 pA (vertical)/20 ms (horizontal). **D**, Patch clamp recording on a parasite exposed with electrical pulses. This second method did not use detergents, which may induce channels in artificial bilayers. This channel is shown in a bursting mode with rapid transitions. Pipette solution, 100 mM CsCl + buffer A; V_p = +30 mV; scale, 2 pA/40 ms. **E**, The channel's current-voltage relationship. Inset shows events at V_p of +40 mV or -70 mV with pipette containing 100 mM CsCl + buffer A (left) or 240 mM glycine + buffer A (right). Scale, 2.8 pA/3 ms. Notice the smaller



amplitudes of channel openings with glycine in the pipette. Main figure shows current-voltage profile with CsCl (filled circles) or glycine (circles) in the pipette. The amplitudes of unambiguous channel events recorded by the two methods described in **A** and **D** were measured by eye. Most patches had multiple channels, making amplitude histograms impractical. Current amplitudes are plotted against the negative of the voltage applied to the pipette. Solid lines, fitted by the least-squares method, demonstrate a decrease in slope conductance and a small change in E_{rev} upon substitution with glycine. Correction for junction potentials (-3 mV for CsCl, -8 mV for glycine) has not been made in the text or in these figures. Channel activity was not seen with isotonic sucrose solutions (used by others to study similar channels²⁹).

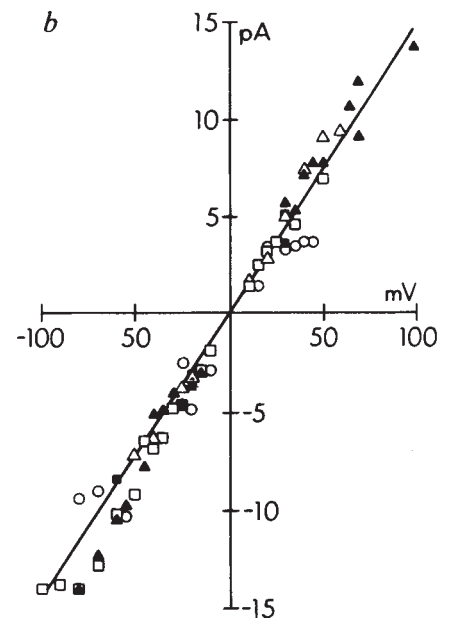
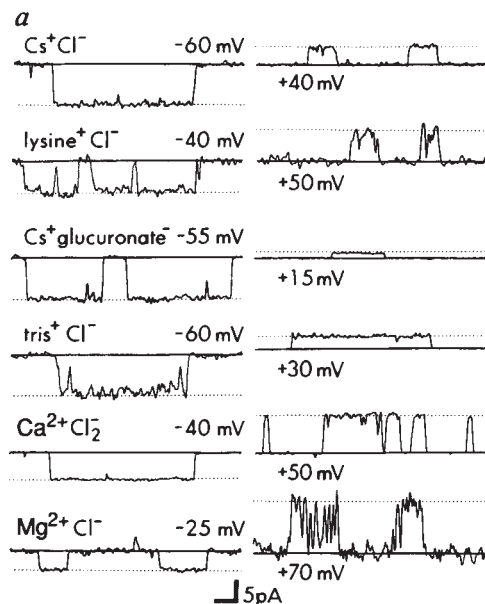
METHODS. *P. falciparum*-infected RBCs were grown in RPMI 1640 culture medium (Gibco) with 10% human serum²⁷ and synchronized using sorbitol lysis²⁸. Single channel currents were recorded from membrane patches on the PVM of trophozoites in the cell-attached configuration⁷. About 50 M Ω pipettes were pulled from Corning 7052 glass capillaries to tip diameters ≤ 1 μ m, heat-polished, filled with filtered pipette solution, and used immediately to ensure gigaohm seals. In **A**, digitonin solutions were made by dissolving dimethylsulphoxide (DMSO) stock solutions in the bath solution (95 mM CaCl $_2$, 5 mM Na-HEPES, pH 7.2) for a final [DMSO] of 0.2%. These differential interference contrast micrographs, taken over ~ 1 min, were printed on a Sony UPS5000 printer.

resting potential, these channels were primarily open: voltage steps away from the resting potential expose the closing transitions of multiple channels. Figure 4 shows that they close within milliseconds when challenged with a -50 mV step, but close only rarely and transiently with a -15 mV step. A similar but less marked inactivation was seen with positive voltage steps (not shown). The physiological significance of this voltage-dependent inactivation is unclear.

We propose that this channel may provide the parasite access to small nutrient molecules in the RBC cytoplasm. Amino acids and monosaccharides, such as lysine and glucuronate (Fig. 2), rapidly enter the RBC cytosol by means of specific transporters

in the RBC membrane^{14,15}. The reported channel could then be used by the parasite to acquire these soluble nutrients and must exist in one of two configurations. In the first, the PVM, studded with these highly permeable channels, would function as a sieve for small molecules entering the space between the PVM and the parasite plasma membrane. Once in this space, nutrients could be internalized by specific transporters in the parasite plasma membrane. Gram-negative bacteria, mitochondria and chloroplasts, which are also surrounded by two membranes and have highly permeable porin channels on their outer membrane, import nutrients by such a mechanism^{16,17}. In the second configuration, the channel would span the parasite plasma mem-

FIG. 2 Permeabilities of various cations and anions. *a*, Inward and outward channel events with six different pipette solutions (predominant ions and V_p values indicated) and *b*, their current-voltage profiles superimposed on the CsCl line of Fig. 1*E* (solid line). Symbols (in *b*) and pipette solutions (mM) are: Cs^+Cl^- , 100 CsCl + buffer A; lysine $^+\text{Cl}^-$ (squares), 120 lysine-Cl, 20 MgCl_2 , 5 TEA-HEPES, pH 7.0; $\text{Cs}^+\text{glucuronate}^-$ (circles), 130 glucuronic acid, 5 MgCl_2 , 7.7 Cs-HEPES, pH 7.0; Tris $^+\text{Cl}^-$ (filled squares), 100 Tris-Cl + buffer A, pH 7.0; $\text{Ca}^{2+}\text{Cl}_2^-$ (triangles), 95 CaCl_2 , 5 Na-HEPES, pH 7.2; $\text{Mg}^{2+}\text{Cl}_2^-$ (filled triangles), 100 MgCl_2 , 0.1 EGTA, 5 TEA-HEPES, pH 7.0. Junction potentials ranged from -8 mV to $+1$ mV (pipette relative to bath) and were ignored. Bath solution was the same as for Fig. 1, except for $\text{Mg}^{2+}\text{Cl}_2^-$ sweeps, when it was 100 MgCl_2 , 0.1 EGTA, 5 TEA-HEPES, pH 7.0, to demonstrate that a high $[\text{Ca}^{2+}]$ is not required for channel activity. The continuous presence of divalent cations in the bath was necessary for gigaohm seal integrity. Horizontal scale in *a* is 3 ms (CsCl), 5 ms (lysine-Cl, CaCl_2 at $+50$ mV), 6 ms (Cs-glucuronate, Tris-Cl at -60 mV, MgCl_2), 8 ms (Tris-Cl at $+30$ mV, CaCl_2 at -40 mV). Although there may be subtle differences in slope conductances with these solutions, each has an E_{rev} indistinguishable from CsCl. Each



major substituted ion thus has a permeability comparable to Cs^+ and Cl^- . K^+ , the predominant cation in RBCs, was not tested because we were unable to form gigaohm seals in its presence.

brane as well as the PVM. This configuration, like that of gap junction channels¹⁸, would allow the direct movement of nutrients from the RBC cytosol to the parasite cytosol. Because the space between the PVM and the parasite plasma membrane may be connected by membranous ducts to the plasma¹⁹, this second configuration would also prevent extracellular leakage of RBC nutrients, which might occur if the channel were to span only the PVM. (The existence of the membranous ducts is highly debated²⁰. If membranous ducts do not exist, neither channel configuration would produce extracellular leakage.) The on-cell patch clamp method used here verifies that the

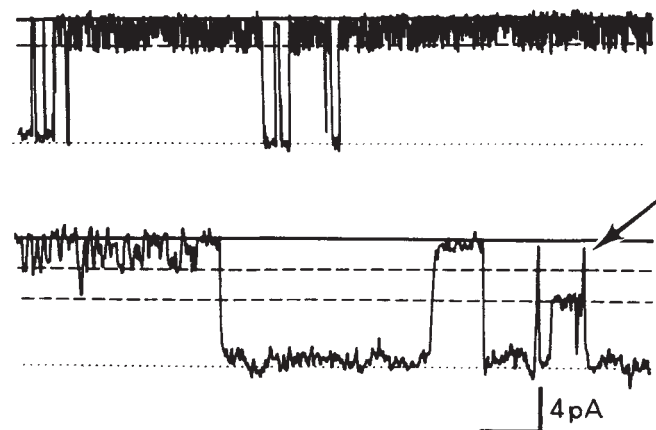


FIG. 3 Subconductance states. Top sweep shows downward openings to the primary conductance described in Figs 1*E* and 2*b* (dotted line) as well as a flickering 25 pS subconductance (dashed line). It is not a separate channel because its flickering disappears during transitions to the primary conductance. The lower sweep, recorded at a faster speed, shows nearly resolved transitions of this subconductance (upper dashed line) as well as a second apparent subconductance level (lower dashed line). Rare transitions suggestive of a conductance greater than the primary 140 pS state were also seen. Pipette solution, 100 CsCl + buffer A; $V_p = -60$ mV; horizontal scale is 70 ms (upper trace), 8 ms (lower trace).

channel allows nutrients to cross the PVM but cannot distinguish between the two channel configurations.

Two observations favour the configuration in which the channel spans only the PVM. First, the intraerythrocytic parasite and the host RBC have dramatically different K^+ and Na^+ contents²¹, suggesting that the two cytosols are not connected by a non-specific channel. Second *P. yoelii* and *P. chabaudi* have negative electrical potentials (E_p) relative to their host RBC^{22,23}. If this non-selective channel crosses both the PVM and the parasite

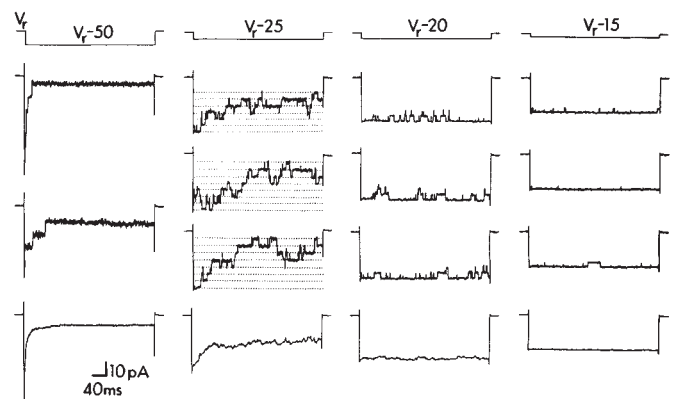


FIG. 4 High density of primarily open channels and voltage dependence. These traces, recorded from a single patch, were obtained by applying 400-ms pulses from V_r (the resting membrane potential, or 0 mV applied to the pipette) to voltages indicated at the top of each column. The bottom trace in each column is an average of 11 sweeps recorded with the same voltage pulse. Pipette solution (mM) was 100 CaCl_2 , 20 MgCl_2 , 5 TEA-HEPES, pH 7.0. There are at least 7 active channels in this patch (see second sweep in second column) which close rapidly with large hyperpolarizing steps (first column) and only rarely as V_r is approached (last column). Using the last column and similar data obtained with small positive voltage steps, we calculate an open probability of $>98\%$ at V_r . Although this open probability does not depend on the pipette solution's composition, an artefactual elevation of open probability by high bath concentrations of divalent cations cannot be ruled out.

plasma membrane, E_p should be substantially diminished by channel activity. Thus, a large E_p in *P. falciparum* would suggest a channel which spans only the PVM.

An inhibitor of this nutrient channel might be an effective antimalarial. Because our technically arduous experimental protocol is not suitable for extensive inhibitor studies, more practical methods will be needed to identify potential inhibitors. We found, however, that 100 μ M chloroquine, a commonly used antimalarial²⁴, did not eliminate channel activity when added to the pipette solution (not shown).

Direct import of soluble RBC nutrients^{1,2}, endocytosis of haemoglobin²⁵, and diffusion through membranous ducts¹⁹ are three mechanisms of nutrient acquisition proposed for the parasite. The channel described here provides a mechanism for the direct import of nutrients and may therefore be an essential adaptation for survival in the RBC. □

Received 26 January; accepted 5 February 1993.

1. Sherman, I. W. *Parasitology* **96**, 557–581 (1988).
2. Tanabe, K. *Parasit. Today* **6**, 225–229 (1990).
3. Trager, W. *J. Protozool.* **18**, 392–399 (1971).
4. Sherman, I. W. *Bull. World Health Org.* **55**, 211–225 (1977).
5. Hansen, B. D., Sleeman, H. K. & Pappas, P. W. *J. Parasit.* **66**, 205–212 (1980).
6. Kanaani, J. & Ginsburg, H. *J. Biol. Chem.* **264**, 3194–3199 (1989).
7. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch.* **391**, 85–100 (1981).
8. Langreth, S. G. *Bull. World Health Org.* **55**, 171–178 (1977).
9. Yamada, K. A. & Sherman, I. W. *Molec. biochem. Parasit.* **3**, 253–264 (1981).
10. Wibo, M., Thines-Sempoux, D., Amar-Costesec, A., Beaufay, H. & Godelaine, D. *J. Cell Biol.* **89**, 456–474 (1981).
11. Dwyer, T. M., Adams, D. J. & Hille, B. *J. gen. Physiol.* **75**, 469–492 (1980).
12. Jap, B. K. & Walian, P. J. *Q. Rev. Biophys.* **24**, 367–403 (1990).
13. Hunter, M. & Giebisch, G. *Nature* **327**, 522–524 (1987).
14. Gero, A. M. & Wood, A. M. *Adv. exp. Med. Biol.* **309A**, 169–172 (1991).
15. Ginsburg, H. *Comp. biochem. Physiol.* **95A**, 31–39 (1990).
16. Colombini, M. *Nature* **279**, 643–645 (1979).
17. Benz, R. *CRC crit. Rev. Biochem.* **19**, 145–190 (1985).
18. Bennett, M. V. L. et al. *Neuron* **6**, 305–320 (1991).
19. Pouvelle, B. et al. *Nature* **353**, 73–75 (1991).
20. Haldar, K. & Uyetake, L. *Molec. biochem. Parasit.* **50**, 161–177 (1992).
21. Lee, P., Ye, Z., Van Dyke, K. & Kirk, R. G. *Am. J. trop. Med. Hyg.* **39**, 157–165 (1988).
22. Mikkelson, R. B., Wallach, E. F. H., Van Doren, E. & Nili, E. A. *Molec. biochem. Parasit.* **21**, 83–92 (1986).
23. Izumo, A., Tanabe, K. & Kato, M. *Comp. biochem. Physiol.* **91B**, 735–739 (1988).
24. Slater, A. F. G. & Cerami, A. *Nature* **355**, 167–169 (1992).
25. Goldberg, D. E. et al. *J. exp. Med.* **173**, 961–969 (1991).
26. Schroeder, J. E., Fischbach, P. S., Zheng, D. & McCleskey, E. W. *Neuron* **6**, 13–20 (1991).
27. Trager, W. & Jensen, J. B. *Science* **193**, 673–675 (1976).
28. Lambros, C. & Vanderberg, J. P. *J. Parasit.* **65**, 418–420 (1979).
29. Franciolini, F. & Wolfgang, N. *J. gen. Physiol.* **90**, 453–478 (1987).

ACKNOWLEDGEMENTS. We thank R. Wilkinson and M. Levy for helping to produce the video micrographs and the electron micrograph, respectively, and S. Dolan, C. Ide, G. Ward and M. Wiser for reviewing the manuscript. This work was supported by the National Heart, Lung and Blood Institute, National Institute of Allergy and Infectious Diseases, National Institute of General Medical Sciences, NIH, and UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases.

Tyrosine kinase activation through the extracellular domains of cytokine receptors

Tomoki Chiba, Yuka Nagata, Mitsuru Machide, Atsushi Kishi, Hiroshi Amanuma, Masahide Sugiyama & Kazuo Todokoro*

Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), 3-1 Koyadai, Tsukuba, Ibaraki 305, Japan

INTERACTION of cytokines with their membrane receptors induces the proliferation and differentiation of a specific lineage of haematopoietic progenitors¹. The molecular mechanism of cytokine receptor-mediated signal transduction is unclear because these receptors do not have tyrosine kinase activity^{1,2}. Interleukin-3 and erythropoietin, however, induce transient tyrosine phosphorylation of a common set of proteins as a growth signal^{3–6}, and interleukin-2

induces phosphorylation of an overlapping but distinct set of proteins^{6–8}. Here we show that chimaeric receptors consisting of the extracellular domains of the erythropoietin receptor and the cytoplasmic domains of the interleukin-2 (or interleukin-3) receptor induce an erythropoietin-dependent tyrosine phosphorylation in interleukin-3-dependent Ba/F3 cells; however, chimaeric receptors composed of the extracellular domains of the interleukin-2 receptor and the cytoplasmic domains of the erythropoietin (or interleukin-3) receptor apparently transmit an interleukin-2-dependent signal. Our results indicate that these cytokines transmit distinct signals for activation of specific tyrosine kinases through the extracellular rather than cytoplasmic domains of the receptors.

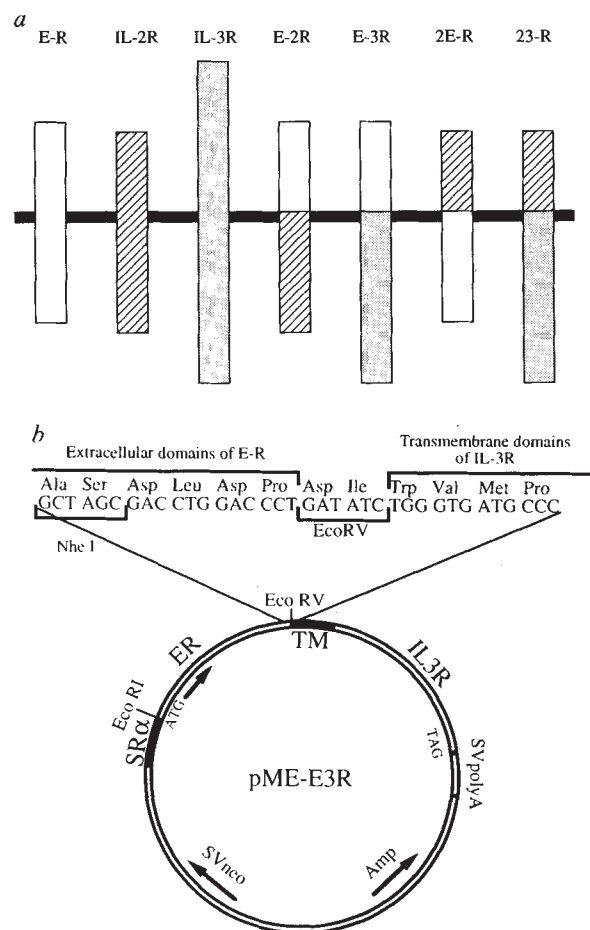


FIG. 1 Structures of chimaeric cytokine receptors. *a*, Schematic representation. Receptors composed of the extracellular domains of E-R and the cytoplasmic domains of IL-2R or IL-3R were designated E-2R and E-3R, respectively. Chimaeras comprising the extracellular domains of IL-2R and the cytoplasmic domains of E-R or IL-3R were designated 2E-R and 23-R, respectively. *b*, Structure of a chimaeric receptor in the expression vector (pME-E3R) and sequence of the joined segment. Only E-3R is shown. The EcoRV site was inserted at the boundary of the transmembrane and extracellular domains such that two amino acids (Asp-Ile) were inserted without frameshift, and the extracellular domains of the receptors were interchanged. METHODS. Cloned cDNAs encoding mouse E-R¹³, human IL-2R¹¹ and mouse IL-3R (Aic2A)¹⁴ were inserted downstream of the SRα promoter in the expression vector pME18, which contained the neomycin-resistance gene with the simian virus-40 promoter¹⁵. The pME-E3R was digested with *NheI* located 13 bp upstream from the transmembrane domains and with *TthIII* located 9 bp inside the transmembrane domains. The double-stranded synthetic oligonucleotides of the *NheI*-*TthIII* segment in which the EcoRV site had been inserted (at the boundary 5'-CTAGCGACCTGGACCTGATATCCTCATCTTGACG-3', 3'-GCTGGACCTGGGACTATAGGAGTAGAAGTGC-5') were replaced with *NheI*-*TthIII* fragment of pME-ER. Similarly, EcoRV sites were inserted at the boundaries of the other cytokine receptor cDNAs, and the chimaeric receptors constructed by interchanging extracellular domains (EcoRI-EcoRV fragments) with each other.

* To whom correspondence should be addressed.

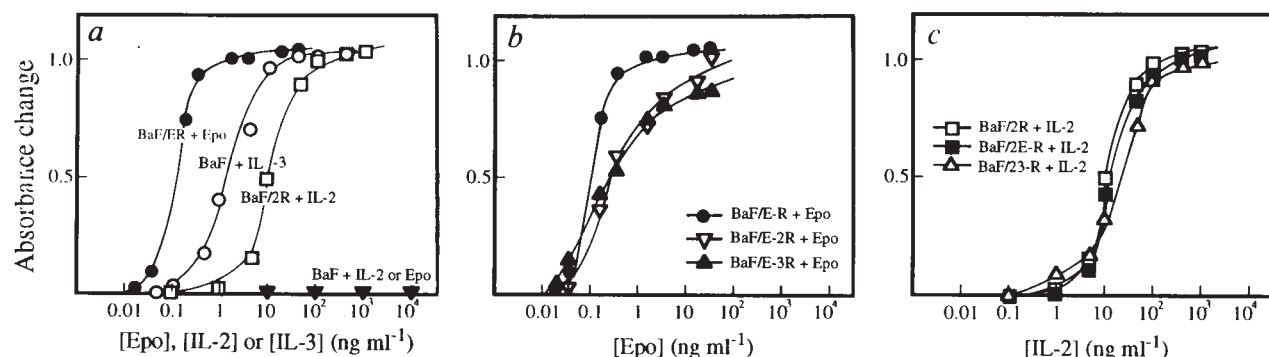


FIG. 2 Ba/F3 transformants expressing E-2R and E-3R chimaeric receptors as well as the wild-type erythropoietin (Epo) receptor grow with Epo in a dose-dependent manner, and the transformants carrying 2E-R and 23-R chimaeras as well as wild-type IL-2R induce a growth signal with IL-2. *a*, Ba/F3 cells expressing wild-type receptors grown with the corresponding cytokines. Ba/F3 cells expressing E-R were stimulated with Epo (●), Ba/F3 cells stimulated with IL-3 (○), Ba/F3 cells expressing IL-2R stimulated with IL-2 (□), and Ba/F3 cells stimulated with IL-2 or Epo (▼). *b*, Chimaeric receptors composed of extracellular domains of E-R transmit a growth signal with Epo. All transformants were stimulated with Epo: Ba/F3 cells expressing E-R (●), Ba/F3 cells expressing E-2R (▽), and Ba/F3 cells expressing E-3R (▲). *c*, Growth of Ba/F3 cells expressing chimaeric receptors comprising the extracellular domains of IL-2R depends on IL-2. All transformants were

induced with IL-2: Ba/F3 cells expressing IL-2R (□), Ba/F3 cells expressing 2E-R (■), and Ba/F3 cells expressing 23-R (△).

METHODS. Transfection of wild-type and chimaeric receptor cDNAs in expression vectors into Ba/F3 cells was by electroporation in 0.8 ml Dulbecco's phosphate-buffered saline containing 2×10^7 cells, 11 mM glucose and 50 μ g of linearized DNA. Cells were pulsed at 600 V and 25 μ F once, and at 250 V and 960 μ F at 4 °C. Stable transformants resistant to antibiotic G418 (1 mg ml^{-1}) were cloned by limiting dilution, and four of each of the transformant clones were analysed. Cell proliferation was measured by a colorimetric assay using 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide¹⁶. Cells (3×10^4) in microtitre plates in 100 μ l culture medium in the presence or absence of various concentrations of IL-2, IL-3 or Epo were cultured at 37 °C for 48 h.

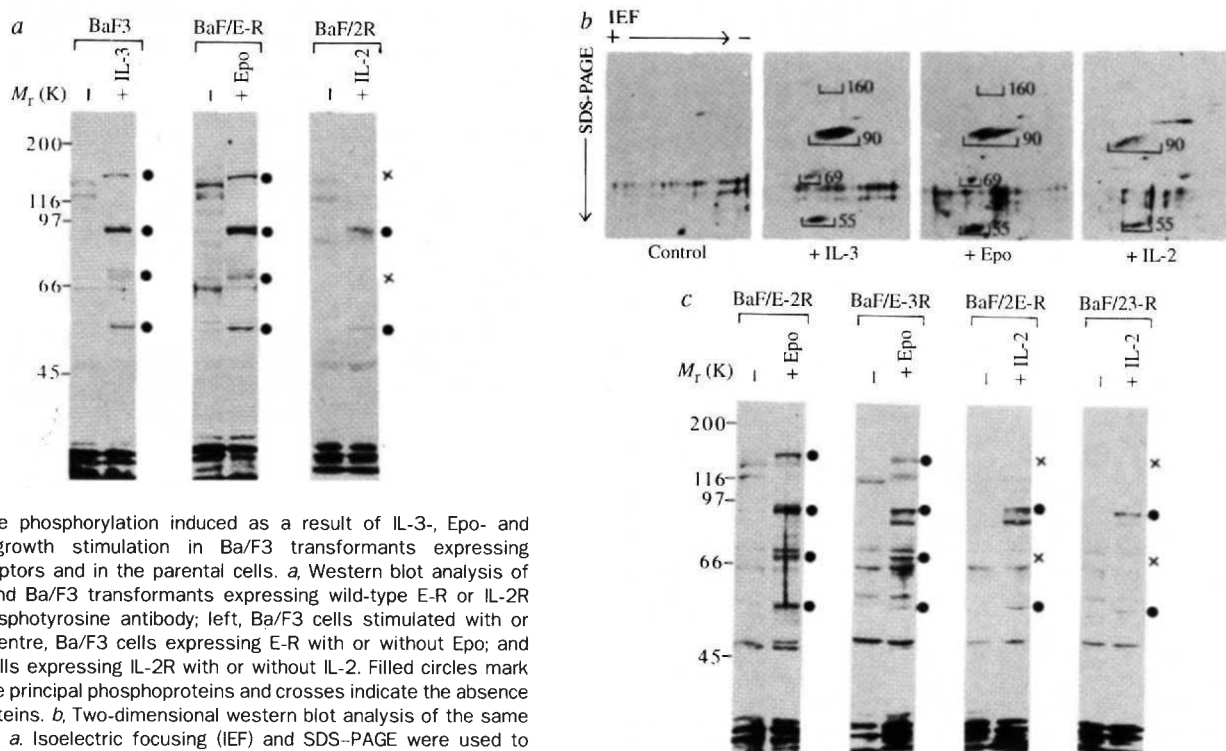


FIG. 3 Tyrosine phosphorylation induced as a result of IL-3-, Epo- and IL-2-induced growth stimulation in Ba/F3 transformants expressing chimaeric receptors and in the parental cells. *a*, Western blot analysis of Ba/F3 cells, and Ba/F3 transformants expressing wild-type E-R or IL-2R using anti-phosphotyrosine antibody; left, Ba/F3 cells stimulated with or without IL-3; centre, Ba/F3 cells expressing E-R with or without Epo; and right, Ba/F3 cells expressing IL-2R with or without IL-2. Filled circles mark positions of the principal phosphoproteins and crosses indicate the absence of phosphoproteins. *b*, Two-dimensional Western blot analysis of the same samples as in *a*. Isoelectric focusing (IEF) and SDS-PAGE were used to resolve phosphotyrosylproteins (numbers show M_r s in thousands). From left to right: unstimulated Ba/F3 cells, Ba/F3 cells stimulated with IL-3, Ba/F3 cells expressing E-R with Epo, Ba/F3 cells expressing IL-2R with IL-2. *c*, Tyrosine phosphorylation of the transformants expressing chimaeric receptors before and after stimulation with the corresponding cytokines. From left to right, Ba/F3 cells expressing E-2R with or without Epo, Ba/F3 cells expressing E-3R with or without Epo, Ba/F3 cells expressing 2E-R with or without IL-2, and Ba/F3 cells expressing 23-R with or without IL-2.

METHODS. Cloned cells grown with IL-3 (Ba/F3 cells), with IL-2 (BaF/2R, BaF/2E-R and BaF/23-R transformants), or with Epo (BaF/E-R, BaF/E-2R and BaF/E-3R transformants) were washed with starvation medium (RPMI 1640 with 1% fetal calf serum) and incubated in the same medium for 10 h. During the last 30 min, 0.7 mM sodium orthovanadate was included in the medium. Starved cells were then stimulated with 250 ng ml^{-1} IL-3 (1.0×10^5 units mg^{-1} ; Genzyme), 11 ng ml^{-1} Epo (2.24×10^5 units mg^{-1} ; Kirin Brewery) or 1 $\mu\text{g ml}^{-1}$ IL-2 (1.0×10^7 units mg^{-1} ; Shionogi Pharmaceutical) for 20 min.

Cells were solubilized in 1% Triton, 10 mM sodium pyrophosphate, 100 mM sodium fluoride, 4 mM EDTA, 2 mM sodium orthovanadate, 1 mM phenylmethylsulphonylfluoride and 50 mM HEPES, pH 7.4. Cell extracts (150 μ g per lane) were applied to a 7.5% SDS-polyacrylamide gel and immunoblotted with monoclonal anti-phosphotyrosine antibody PY20 (ICN) as described³. Phosphoproteins were detected by an ECL system (Amersham). Two-dimensional gel electrophoresis was according to ref. 17 with some modifications. Samples were acetone-precipitated and redissolved in 9 M urea, 4% Norit-P40, with 2% ampholyte (pI values, 3–10; Pharmacia) and 2% β -mercaptoethanol. Samples were loaded onto a 1.5-mm tube gel and isoelectrically focused at 400 V for 16 h. After removal of gels from the tubes, they were equilibrated with 3% SDS, 5% β -mercaptoethanol and 125 mM Tris, pH 6.8, and then electrophoresed on a 7.5% SDS-polyacrylamide gel. Western blots of the two-dimensional gels were then analysed using anti-phosphotyrosine antibody.

To clarify how cytokine receptors having no tyrosine kinase activity can transmit a growth or differentiation signal to specific tyrosine kinases, we constructed various chimaeric receptors (Fig. 1a) and studied tyrosine phosphorylation induced by chimaeric receptor-mediated growth signals. Chimaeric receptors were prepared by inserting an *EcoRV* restriction enzyme digestion site at the boundary of the transmembrane and extracellular domains, followed by interchange of the extracellular domains (Fig. 1b). The Pro-B cell line Ba/F3 (ref. 9) is dependent on the interleukin (IL) IL-3, but not on erythropoietin or interleukin-2, for cell growth (Fig. 2a). The exogenous wild-type erythropoietin receptor (E-R), the IL-2 receptor β -chain (IL-2R), or the chimaeric receptors shown in Fig. 1 (E-2R, E-3R, 2E-R and 23-R) were transfected and expressed in Ba/F3 cells. Cloned stable transformants were then assayed for cell proliferation using each cytokine. Scatchard analysis of the binding of iodinated erythropoietin or IL-2 to stable transformants confirmed that wild-type and chimaeric receptors are expressed on each cloned transformant with a frequency of $\sim 10^3$ per cell and with affinities as high as those of the intact receptors^{10,11} (data not shown). As shown in Fig. 2a and b, Ba/F3 cells expressing chimaeric receptors composed of the extracellular domains of E-R (E-2R and E-3R) proliferated in response to erythropoietin in a dose-dependent manner, as did Ba/F3 cells expressing wild-type E-R. Similarly, growth of transformants expressed chimaeric receptors comprising the extracellular domains of IL-2R (2E-R and 23-R) and those expressing wild-type IL-2R was dependent on the concentration of IL-2 (Fig. 2a and c).

Phosphorylation of tyrosine residues upon growth stimulation was then investigated in Ba/F3 transformants expressing exogenous wild-type or chimaeric receptors, as well as in the parental cells. Western blot analysis with antiphosphotyrosine antibody showed that erythropoietin and IL-3 induced tyrosine phosphorylation of the same major proteins (M_r s 55, 69, 90 and 160K) in Ba/F3 cells expressing E-R and in the parental cells (Fig. 3a). Some minor proteins specific for IL-3 stimulation were also seen. IL-2 induced tyrosine phosphorylation on only two of the major proteins (M_r s 55 and 90K, but not those of 69 and 160K) in Ba/F3 transformants expressing IL-2R (Fig. 3a). These results are consistent with data from IL-2-, IL-3- or erythropoietin-dependent mouse cell lines³⁻⁸. We ran two-dimensional western blots to distinguish the phosphoproteins more clearly. The results showed that the major phosphoproteins induced by IL-3 and erythropoietin are identical, and that IL-2 induces the phosphorylation of the 55K and 90K proteins but not of the 69K and 160K proteins, indicating that IL-2 and IL-3 (or erythropoietin) trigger tyrosine phosphorylation of an overlapping but distinct set of proteins. These results are consistent with recent reports^{7,12} that IL-2 and IL-3 induce distinct but overlapping responses.

Figure 3c shows the phosphorylation patterns of transformants expressing the E-2R, E-3R, 2E-R and 23-R chimaeric receptors. Chimaeric receptors E-2R and E-3R transmit the same signal as the wild-type E-R (Fig. 3a, b), even though the cytoplasmic domains of these receptors are composed of IL-2R or IL-3R. 2E-R and 23-R chimaeric receptors, in contrast, induce IL-2-dependent (but not erythropoietin- or IL-3-dependent) tyrosine phosphorylation (Fig. 3b), although the cytoplasmic domains of the receptors consist of E-R and IL-3R. Thus, tyrosine phosphorylation depends on the constituents of the extracellular rather than cytoplasmic domains of the receptors. These results indicate that signals for activation of specific tyrosine kinases are mediated through the extracellular domains of receptor molecules. Therefore the specific interaction of another membrane component or signal transducer with the extracellular domains of the cytokine receptors is critical for transmission of a signal to activate specific tyrosine kinases, although an influence of the receptor cytoplasmic domains on tyrosine phosphorylation cannot be completely ruled out. □

Received 19 October 1992; accepted 1 February 1993.

1. Arai, K. *et al.* *Rev. Biochem.* **59**, 783-836 (1990).
2. Miyajima, A., Kitamura, T., Harada, N., Yokota, T. & Arai, K. *Rev. Immun.* **10**, 295-331 (1991).
3. Kuramochi, S., Chiba, T., Amanuma, H., Tojo, A. & Todokoro, K. *Biochem. biophys. Res. Commun.* **181**, 1103-1109 (1991).
4. Quelle, F. W. & Wojchowski, D. M. *J. Biol. Chem.* **266**, 609-614 (1991).
5. Isfort, R., Huhn, R. D., Frackelton, A. R. & Ihle, J. N. *J. Biol. Chem.* **263**, 19203-19209 (1988).
6. Morla, A. O., Schreurs, J., Miyajima, A. & Wang, J. Y. *J. Molec. cell. Biol.* **8**, 2214-2218 (1988).
7. Sabe, H. *et al.* *Intl. Immun.* **3**, 1137-1148 (1991).
8. Saltzman, E. M., Lufwinsky, S. M. & Casnellie, J. E. *J. Biol. Chem.* **264**, 19979-19983 (1989).
9. Palacios, R. & Steinmetz, M. *Cell* **41**, 727-734 (1985).
10. Todokoro, K., Kanazawa, S., Amanuma, H. & Ikawa, Y. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4126-4130 (1987).
11. Hatakeyama, M. *et al.* *Science* **244**, 551-556 (1989).
12. Otani, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **89**, 2789-2793 (1992).
13. Kuramochi, S., Ikawa, Y. & Todokoro, K. *J. molec. Biol.* **216**, 567-575 (1990).
14. Itoh, N. *et al.* *Science* **247**, 324-327 (1990).
15. Kitamura, T. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 5082-5086 (1991).
16. Chiba, T., Amanuma, H. & Todokoro, K. *Biochem. biophys. Res. Commun.* **184**, 485-490 (1992).
17. O'Farrell, P. H. *J. Biol. Chem.* **250**, 4007-4021 (1975).

ACKNOWLEDGEMENTS. We thank A. Miyajima for IL-2R and IL-3R cDNAs.

Budding from Golgi membranes requires the coatomer complex of non-clathrin coat proteins

Lelio Orci*, David J. Palmer†, Mariella Ravazzola*, Alain Perrelet*, Mylène Amherdt* & James E. Rothman†

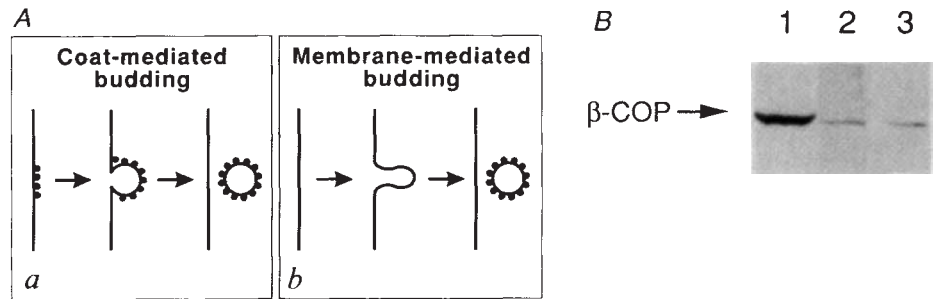
* Institute of Histology and Embryology, University of Geneva Medical School, 1 rue Michel Servet, 1211 Geneva 4, Switzerland

† Program in Cellular Biochemistry and Biophysics, Sloan-Kettering Institute, 1275 York Avenue, New York, New York 10021, USA

DO the coats on vesicles budded from the Golgi apparatus actually cause the budding, or do they simply coat buds (Fig. 1)? One view (the membrane-mediated budding hypothesis¹) is that budding is an intrinsic property of Golgi membranes not requiring extrinsic coat proteins. Assembly of coats from dispersed subunits is superimposed upon the intrinsic budding process and is proposed to convert the tips of tubules into vesicles. The alternative view (the coat-mediated budding hypothesis¹) is that coat formation provides the essential driving force for budding. The membrane-mediated budding hypothesis was inspired by the microtubule-dependent extension of apparently uncoated, 90-nm-diameter membrane tubules from the Golgi apparatus² and other organelles³⁻⁵ *in vivo* after treatment with brefeldin A, a drug that inhibits the assembly of coat proteins onto Golgi membranes⁶⁻⁹. This hypothesis predicts that tubules will be extended when coat proteins are unavailable to convert tubule-derived membrane into vesicles. Here we use a cell-free system in which coated vesicles are formed from Golgi cisternae to show that, on the contrary, when budding diminishes as a result of immunodepletion of coat protein pools, tubules are not formed at the expense of vesicles. We conclude that coat proteins are required for budding from Golgi membranes.

When Golgi membranes are incubated with cytosol and ATP, intercisternal transport is reconstituted¹⁰, and non-clathrin coated vesicles are formed¹¹. On addition of GTP- γ S, transport is blocked and the coated vesicles accumulate¹², allowing purification of the vesicles^{13,14} and the identification of the principal coat proteins (COPs). The major component of the coat is coatomer¹⁵, a complex of seven polypeptides with M_r s ranging from 20,000 to 160,000 which remain associated in the cytosol, but it also contains stoichiometric amounts of a small GTP-binding protein, ADP-ribosylation factor (ARF; refs 16, 17). As Golgi membrane themselves contain only trace quantities of coatmer and ARF (refs 17, 18), these components of the vesicles formed in the cell-free system must come from the cytosol.

FIG. 1 *Aa*, Coat-mediated and *b*, membrane-mediated budding processes, as defined in ref. 1. "On the left is shown the coat-mediated budding model in which cytosolic coat proteins assemble onto the membrane and undergo a structural transition which provides the mechanochemical force to induce budding. In contrast, membrane-mediated budding (shown on the right) occurs in the absence of assembly of cytosolic coat proteins. The bud can serve as the site of assembly of coat proteins and/or continue to grow as an uncoated tubule." *B*, Western blot analysis of coatomer-depleted bovine brain cytosol. Bovine brain cytosol was immunodepleted of coatomer using the monoclonal antibody CM1A10 and analysed by western blotting using the monoclonal antibody against β -COP, M3A5 (ref. 28). Lane 1, bovine brain cytosol, 18 μ g protein; lane 2, coatomer-depleted bovine brain cytosol, 18 μ g protein; lane 3, bovine brain cytosol, 2.25 μ g protein.



METHODS. CM1A10, a mouse monoclonal IgG, was bound to protein G-agarose by incubation of CM1A10 ascites (7.4 mg IgG per ml) with 1/6 vol of protein G-agarose (Calbiochem) for 16 h at 4 °C with mixing. The agarose beads were then washed four times with 10 vol 0.2 M sodium borate, pH 9.0, at 4 °C, and bound CM1A10 coupled to protein G-agarose by incubation with 20 mM dimethylsuberimide in 0.2 M sodium borate, pH 9.0, at room temperature for 30 min with mixing. The crosslinking reaction was terminated

by extensive washing with 0.2 M ethanolamine, pH 8.0. The CM1A10-protein G-agarose (8.8 mg IgG per ml of 50% suspension) was stored in phosphate-buffered saline at 4 °C. Bovine brain cytosol (18 mg of protein per ml) was prepared as described¹³ and was immunodepleted of coatomer by sequential incubations with CM1A10-protein G-agarose (13 μ g IgG per mg cytosol protein per incubation) for 6 h, 6 h and 12 h at 4 °C with mixing. For western blot analysis, samples were fractionated by SDS-PAGE using a 5% gel under reducing conditions. After electrophoresis, proteins were electroblotted onto nitrocellulose and reacted with M3A5 and peroxidase-conjugated anti-mouse IgG. Peroxidase-labelling was detected by chemiluminescence using ECL reagent (Amersham).

To deplete cytosol of coatomer, we have used a monoclonal IgG antibody directed against the coatomer (monoclonal antibody CM1A10) to precipitate the complex of COP proteins (including β -COP)³⁰ (Fig. 1*B*). About 85% of the β -COP was immunodepleted; further immunoadsorptions did not remove additional β -COP (data not shown). Removal of coatomer (Table 1) reduced the number of vesicles plus buds bearing coat from 25.3 ± 1.3 per μm^2 (Fig. 2*d*) to 6.2 ± 0.7 per μm^2 (Fig. 2*e*). When more than native amounts of coatomer were added to depleted cytosol, approximately 40% more β -COP become bound to Golgi membranes (Table 1, columns A and E) than with untreated cytosol, and the number of coated vesicles and buds rose to 39.3 ± 1.5 per μm^2 . No significant coated structures (0.3 ± 0.1 per μm^2) formed when coatomer was added alone, without immunodepleted cytosol, implying that cytoplasmic factors other than coatomer are also required to form coated vesicles. Examples of the vesicles and buds formed are shown in Fig. 2*a-c*. Partial removal of coatomer by immunodepletion does not appreciably affect the extent of cell-free transport

within the Golgi (data not shown). More complete removal of coatomer, by other methods, has effects fully consistent with transport mediated by coated vesicles, and will be described elsewhere.

Most of the membrane-bound β -COP and ARF proteins (Table 1) were incorporated in coats. Depletion of coatomer did not substantially reduce the amount of ARF bound to Golgi membranes measured by immunogold labelling (Table 1, columns G and H) or western blotting (data not shown). This is expected because ARF protein can bind to Golgi membranes without coatomer^{19,20}. Depletion of coatomer leaves 81% of the bound ARF outside coats (that is, bound to cisternal membranes), but only 34% remains there when coatomer is added back. This suggests that ARF is recruited to nascent coated vesicles by coatomer, consistent with its role as a coat protein¹⁷ and in coat assembly^{19,30}.

To ascertain whether budding continues when coatomer is removed, producing uncoated tubules and uncoated buds, we examined consecutive serial sections (Fig. 2, and Table 2). As

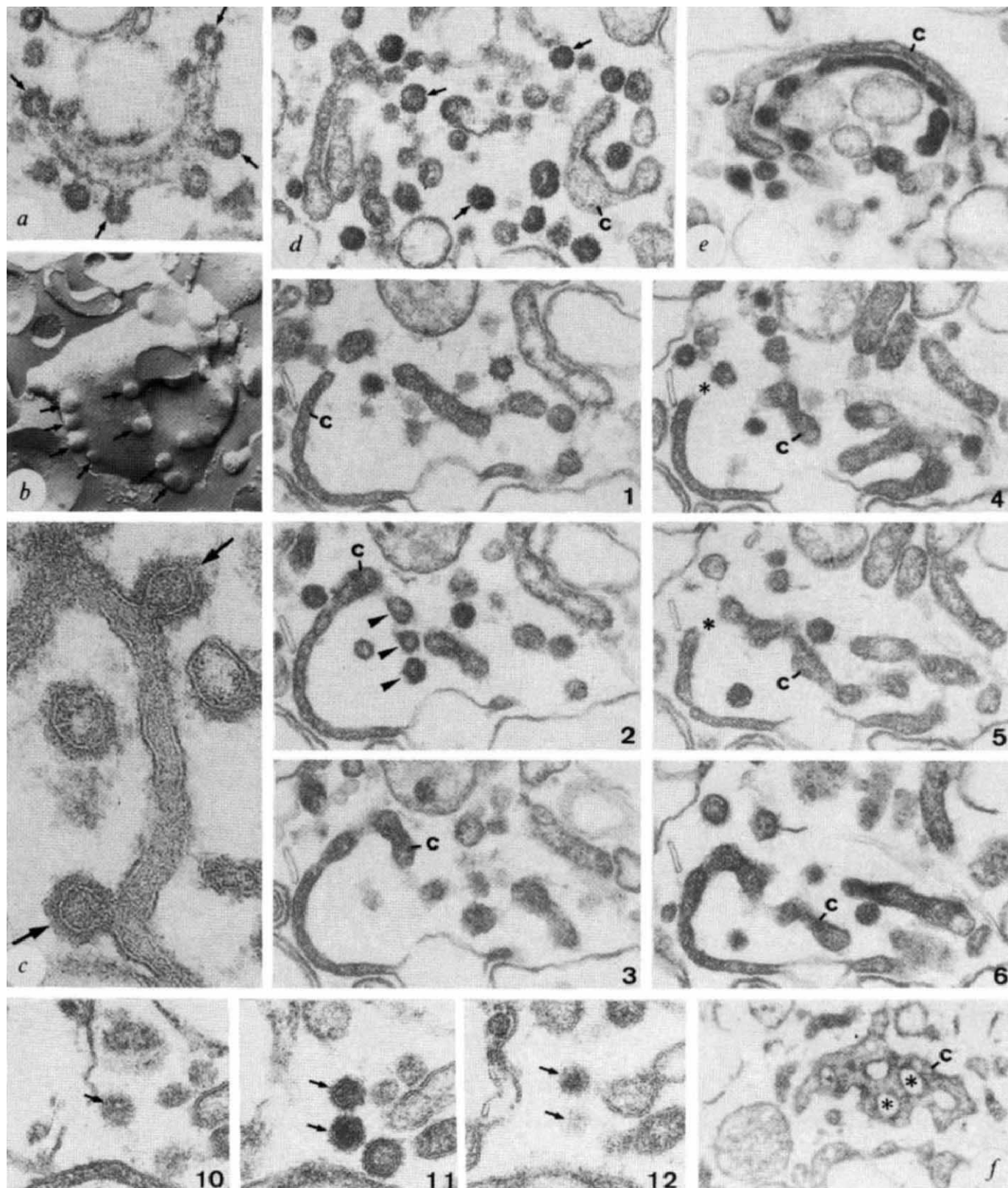
TABLE 1 Requirements for production of COP-coated vesicles in presence of GTP- γ S

	A	B	C	D	E	F	G	H
	β -COP bound to Golgi membranes (arbitrary units)	Uncoated circular profiles (number per μm^2)	Coated vesicles (number per μm^2)	Coated buds (number per μm^2)	Density of β -COP in Golgi areas (number of gold particles per μm^2)	% β -COP labelling associated with coated vesicles plus coated buds	Density of ARF in Golgi areas (number of gold particles per μm^2)	% ARF labelling associated with coated vesicles and coated buds
(1) Golgi + cytosol	6.1	1.2 ± 0.2	22.1 ± 1.3	3.2 ± 0.3	51 ± 3	85 ± 1	40 ± 2	65 ± 3
(2) Golgi + depleted cytosol	1.4	0.9 ± 0.2	4.2 ± 0.7	2.0 ± 0.3	13 ± 2	31 ± 6	33 ± 2	19 ± 2
(3) Golgi + coatomer	[1]	1.0 ± 0.3	0.2 ± 0.1	0.1 ± 0.06	7 ± 2	11 ± 4	11 ± 2	6 ± 4
(4) Golgi + depleted cytosol + coatomer	8.4	0.5 ± 0.2	35.7 ± 1.4	3.6 ± 0.4	68 ± 3	91 ± 1	49 ± 3	66 ± 2

Incubations were as described in the legend to Fig. 2. For electron microscopy, membranes were collected by centrifugation and fixed with glutaraldehyde⁸. Pellets were either contrasted with tannic acid and embedded in Epon (columns B, C and D), or cryosectioned (E-H). Quantitation has been described^{8,11}. Dilution of the anti- β -COP antibody (affinity-purified anti-EAGE) was 1:40; dilution of the anti-ARF antibody (raised against recombinant ARF1 protein) was 1:100. All quantitation was done in a double-blind manner with a code that was not broken until quantitation was completed. Values are means $\pm$ s.e.m. based on a minimum sampling at twenty Golgi from two different experiments. For western blot analysis of β -COP bound to Golgi, membranes were recovered from 50 μ l reaction mixtures (Fig. 2 legend) by overlaying on 150 μ l 15% (w/w) sucrose cushions and centrifugation, using 0.5 ml tubes that had been previously incubated with BSA (10 mg ml⁻¹) for 30 min and washed with water, or 30 min in a microcentrifuge at 4 °C. β -COP in the pellets was estimated by western blotting (using M3A5, as in Fig. 1) and densitometry. The values in column A were normalized to the value in (3).

tubules produced by brefeldin A treatment *in vivo* are about 90 nm in diameter¹, we defined a tubule as a membrane-bound cylindrical structure with a diameter ≤ 90 nm and a length more than twice its diameter. The tubules envisioned in the membrane-mediated budding hypothesis are of two possible types: closed at one end (forming from a cisterna at the other end) or closed at both ends, having once budded but now existing separately from a cisterna. All such structures examined (out

of 179) in serial sections of Golgi fractions incubated with immunodepleted cytosol with or without coatomer extended to at least three consecutive sections, and were thus in fact profiles of cisternae; none were tubules. Cross-sections of the disk-shaped Golgi cisternae appear in numerous consecutive sections, unlike longitudinal sections of tubules, which should be limited to at most two (given the 70–80 nm thickness of each section).



Any tubules rising vertically through a series of sections will appear as circular profiles of ≤ 90 nm diameter in any one section. As they must by definition be at least twice as long as they are wide, a circular profile due to a tubule will therefore extend into at least three consecutive sections, opening into cisternae at no more than one end. A circular profile was classed as a vesicle if it did not extend beyond two serial sections, and as a bud if it did not extend beyond two sections and if the circular profile opened into a cisterna in an adjoining section. In extracts depleted of coatamer, there were only 0.9 ± 0.2 uncoated circular profiles per μm^2 (Table 1), of which 68% were classed as vesicles and none were classed as tubules or buds (Table 2) (105 examined). With added coatamer, there were 0.5 ± 0.2 per μm^2 (Table 1) (so none of statistical significance), of which 85% were vesicles and none were tubules or buds (Table 2) (100 examined). All the non-vesicular circular profiles merged in plane with a sheet-like profile at each end of the cylinder, indicating that they represented cylindrical segments within a fenestrated Golgi cisterna. Such fenestrations (Fig. 2f, asterisks), in which perforations are surrounded by cylindrical shaped segments of a cisterna, are a classic feature of Golgi structure. The membrane-mediated budding hypothesis predicts that upon removing coatamer, for every coated vesicle lost one tubule will be gained. The fenestrated cisternae cannot be due to the predicted tubules that might have repeatedly fused to yield the fenestrated topology, because the number of coated

TABLE 2 Analysis of uncoated circular profiles (diameter ≤ 90 nm) from consecutive serial sections

	Buds	Tubules	Vesicles	Cylindrical segments of Golgi cisternae in fenestrated regions
(1) Coatamer-depleted cytosol plus coatamer	ND	ND	0.25 ± 0.03	0.04 ± 0.01
(2) Coatamer-depleted cytosol	ND	ND	0.32 ± 0.06	0.14 ± 0.05

See text for definitions of categories. For (1), 11 Golgi areas were evaluated; for (2), 8 Golgi areas were evaluated. Numbers are means $\pm$ s.e.m. per μm^2 of photograph area. ND, none detected.

vesicles and buds decreases by 33 per μm^2 on depleting coatamer (Table 1), whereas the number of cylindrical elements in fenestrated regions rises by only 0.10 per μm^2 (Table 2).

These observations indicate that less than 1 tubule is formed per 1,000 vesicles even by extracts depleted of coat proteins. The fixation conditions used were similar to those that preserve tubules in whole cells, so the possibility that existing tubules were not preserved seems remote. The *in vitro* system we have used may not mimic the *in vivo* morphological behaviour of the Golgi in all respects. It does bud coated vesicles, however, and according to the membrane-mediated budding hypothesis, the same process gives rise to both tubules and coated vesicles, so it can therefore be used to test the hypothesis.

Tubules can apparently be formed *in vivo* from Golgi and from endosomes²¹. Rather than forming by membrane-mediated budding, they may instead be produced by motor-dependent extension of membrane along microtubules²², thus explaining the requirement for microtubules²². Such microtubules and motors may be absent or inoperative in our cell-free system. This mechanism is not a form of membrane-mediated budding, as it entails deformation of the membrane by an extrinsic system (microtubules).

Coatamer depletion therefore reduces budding from Golgi cisternae without inducing tubule formation, as predicted by the coat-mediated budding hypothesis. The co-assembly of ARF and coatamer into coats presumably drives deformation of the membrane into a vesicle, as has recently been demonstrated for clathrin coats²³⁻²⁵. A role in budding does not rule out other roles for the coat, such as selecting vesicle content²³ and preventing premature vesicle fusion. Potentially fusogenic proteins may be rendered inactive by the coat until it is peeled off at the target membrane²⁶, so that initiation of fusion can be coupled to sequential budding and targeting events. These processes may be uncoupled by brefeldin A⁸.

While this letter was in press, a claim appeared that tubules form extensively from Golgi cisternae upon incubation under certain conditions²⁷. As serial section analysis was not performed and results from negative staining were not quantitated, these putative 'tubules' could well have been images of cisternae and fenestrated cisternae, similar to the structures revealed here by serial sectioning. □

Received 16 November 1992; accepted 4 February 1993.

- Klausner, R. D., Donaldson, J. G. & Lippincott-Schwartz, J. *J. Cell Biol.* **116**, 1071-1080 (1992).
- Lippincott-Schwartz, J. *et al. Cell* **60**, 821-836 (1990).
- Hunziker, W., Whitney, J. A. & Mellman, I. *Cell* **67**, 617-628 (1991).
- Lippincott-Schwartz, J. *et al. Cell* **67**, 601-616 (1991).
- Wood, S. A., Park, J. E. & Brown, W. J. *Cell* **67**, 691-600 (1991).
- Donaldson, J. G. *et al. J. Cell Biol.* **111**, 2295-2306 (1990).
- Duden, R. *et al. Cell* **64**, 649-665 (1991).
- Orci, L. *et al. Cell* **64**, 1183-1195 (1991).
- Robinson, M. S. & Kreis, T. E. *Cell* **69**, 129-138 (1992).
- Balch, W. E. *et al. Cell* **39**, 405-416 (1984).

FIG. 2 Micrographs of Golgi areas in membrane pellets. All incubations contained ATP and ATP-regenerating system, and GTP- γ S. For electron microscopy, membranes were collected by centrifugation and fixed with glutaraldehyde, contrasted with tannic acid and quantitated as described previously⁸. Quantitation is presented in Tables 1 and 2. Precise incubation conditions were as follows: rabbit liver Golgi membranes ($152 \mu\text{g ml}^{-1}$, prepared as described in ref. 13) were incubated for 20 min at 37°C with $20 \mu\text{M}$ GTP- γ S, 25 mM HEPES/KOH, pH 7.2, 25 mM KCl, 0.2 M sucrose, 2.5 mM magnesium acetate, $50 \mu\text{M}$ ATP, 2 mM creatine phosphate, 8 IU ml^{-1} creatine kinase, $250 \mu\text{M}$ UTP and 1 mg ml^{-1} soybean trypsin inhibitor in the presence or absence of coatamer ($8 \mu\text{g ml}^{-1}$; purified as described in ref. 29, using gel filtration as the final purification step), and bovine brain cytosol or coatamer-depleted bovine brain cytosol (1.8 mg ml^{-1} each, prepared as described in the legend to Fig. 1). This plate illustrates the Golgi compartments which were quantitated (a-f), and a series of consecutive thin sections to show the complex geometry of the cisternal component of the Golgi (sections 1-6), as well as the three-dimensional extension of coated vesicles (sections 10-12). a, Thin section of a Golgi cisterna giving rise to several coated buds (arrowed); b, freeze-fracture replica showing the emergence of buds (arrowed) from the surface of a cisternal membrane. Note the emergence of buds in discrete areas of the membrane; c, thin section at high magnification showing part of a cisterna giving rise to two coated buds (arrowed). Such images were scored as 'coated buds' in the quantitative evaluation (Table 1). Non-coated equivalents of these buds were never seen. The coated bud is roughly the size of a free coated vesicle ($\sim 75 \text{ nm}$). d, Field of a Golgi fraction after incubation in the presence of coatamer plus coatamer-depleted cytosol and GTP- γ S. The abundance of coated vesicles (arrowed) is apparent. The cisternal component of the Golgi is indicated as 'c'. Compare, with e, the image showing a Golgi fraction after incubation without coatamer, but in presence of coatamer-depleted cytosol and GTP- γ S. Sections 1-6, consecutive thin sections of a Golgi incubated as for e; the sequence reveals that small (~ 60 - 80 nm) uncoated circular profiles (arrowheads in section 2), which could be interpreted as transverse views of tubules, are in fact in-plane components of the complex geometrical structure of a Golgi cisterna (c) made up of communicating, sheet-like and cylinder-like elements. Sheets are perforated by fenestrations (asterisks in sections 4 and 5, and in f); the fenestrated appearance of a cisterna in a rare example of a grazing section is seen in f. Sections 10-12: Unlike sheet-like or cylindrical elements of fenestrated cisternae that can be followed in several consecutive sections, coated (arrows in section 11) and uncoated vesicles are clearly visible with their limiting membrane in only one section. Adjacent sections (10 or 12) show only 'remnants' (grazing views of the vesicle). Sections 10-12 belong to a series different from that containing sections 1-6. Magnifications: a, $78,100\times$; b, $42,000\times$; c, $169,900\times$; d, $52,600\times$; e, $52,600\times$; f, $47,700\times$; sections 1-6, $59,100\times$; sections 10-11, $82,200\times$. a-c, Golgi fraction from CHO cells; d-f and sections 1-6, 10 and 11, Golgi fraction from rabbit liver.

11. Orci, L., Glick, B. S. & Rothman, J. E. *Cell* **46**, 171-184 (1986).
12. Melancon, P. et al. *Cell* **51**, 1053-1062 (1987).
13. Malhotra, V. et al. *Cell* **58**, 329-336 (1989).
14. Serafini, T. et al. *Nature* **349**, 215-220 (1991).
15. Waters, M. G., Serafini, T. & Rothman, J. E. *Nature* **349**, 248-251 (1991).
16. Kahn, R. A. et al. *J. Biol. Chem.* **266**, 2606-2614 (1991).
17. Serafini, T. et al. *Cell* **67**, 239-253 (1991).
18. Donaldson, J. G. et al. *Science* **254**, 1197-1199 (1991).
19. Donaldson, J. G. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 6408-6412 (1992).
20. Helms, J. B. & Rothman, J. E. *Nature* **360**, 352-354 (1992).
21. Mellman, I. & Simons, K. *Cell* **68**, 829-840 (1992).
22. Pelham, H. R. B. *Cell* **67**, 449-451 (1991).
23. Pearse, B. M. F. & Bretscher, M. S. A. *Rev. Biochem.* **50**, 85-101 (1981).

24. Moore, M. S., Mahaffey, D. T., Brodsky, F. M. & Anderson, R. G. W. *Science* **236**, 558-563 (1987).
25. Smythe, E., Pypaert, M., Lucocq, J. & Warren, G. *J. Cell Biol.* **108**, 843-853 (1989).
26. Orci, L. et al. *Cell* **56**, 357-368 (1989).
27. Cluett, E. B., Wood, S. A., Banta, M. & Brown, W. J. *J. Cell Biol.* **120**, 15-24 (1993).
28. Allan, V. J. & Kreis, T. E. *J. Cell Biol.* **103**, 2229-2239 (1986).
29. Waters, M. G., Beckers, C. J. M. & Rothman, J. E. *Meth. Enzym.* **219**, 331 (1992).
30. Palmer, D. J., Orci, L. & Rothman, J. E. *J. Biol. Chem.* (in the press).

ACKNOWLEDGEMENTS. We thank T. Kreis for the anti-EAGE antibody used for β -COP immunocytochemistry, and B. Helms for anti-ARF1 antibody. This work was supported by an NIH grant to J.E.R., a Swiss National Science Foundation grant to L.O., and a grant from the Human Frontier Science Organization. D.P. was the recipient of an NIH Fogarty International Research Fellowship.

Defective mismatch binding and a mutator phenotype in cells tolerant to DNA damage

P. Branch, G. Aquilina*, M. Bignami* & P. Karran†

Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Herts EN6 3LD, UK

* Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy

ACQUIRED resistance to alkylating agents such as *N*-methyl-*N*-nitrosourea or *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine results from the ability to tolerate the potentially cytotoxic methylated base *O*⁶-methylguanine (*m*⁶-G) in DNA. In the absence of repair by demethylation *in situ*, *m*⁶-G is probably lethal through its inappropriate processing by the cell¹. DNA mismatch correction is an attractive candidate for the processing function because although it is replicated, *m*⁶-G has no perfect complementary base. Thus, *m*⁶-G in DNA might provoke abortive mismatch repair and tolerance could subsequently arise through loss of a mismatch repair pathway^{2,3}. Mismatch correction helps maintain genomic fidelity by removing misincorporated bases and deaminated 5-methylcytosine from DNA, and its loss by mutation confers a mutator phenotype on *Escherichia coli*^{4,5}. Here we describe human and hamster cell lines that are tolerant to *N*-methyl-*N*-nitrosourea and are defective in a DNA mismatch binding activity. The loss of this activity, which acts on G-T mispairs, confers a mutator phenotype.

RajiMex⁻ cells lack the methyltransferase that normally repairs *m*⁶-G in DNA⁶ and are consequently highly sensitive to *N*-methyl-*N*-nitrosourea (MNU). Growth is abolished by treatment with 0.01 mM MNU, whereas proliferation of the related MNU-tolerant RajiF12 cells is unaffected by concentrations of up to 0.3 mM (Fig. 1a, b). The Chinese hamster ovary cell line CHOMT⁺ expresses a low level of *m*⁶-G-DNA methyltransferase. The MNU-resistant derivative, clone B (Fig. 1c), retains this low level of methyltransferase activity but is additionally tolerant to MNU. Both clone B and RajiF12 display resistance to the narrow range of methylating agents that is characteristic of the tolerant phenotype (ref. 7, and our unpublished results). The tolerant phenotype is normally stable in the absence of selective pressure. During routine maintenance over a period of months, however, a culture of clone B lost resistance to MNU. A subclone of this spontaneous revertant population, designated BS11, showed MNU-sensitivity comparable to CHOMT⁺ (ref. 8).

Extracts of mammalian cells contain proteins that bind specifically to single DNA mispairs in oligonucleotides^{9,10}. Two activities are distinguishable by the migration characteristics of complexes formed with oligonucleotides containing different mismatches. One complex is formed with G-T mismatches, whereas a more rapidly migrating complex is formed with A-C and some pyrimidine-pyrimidine mispairs. Both binding activities were found in extracts of the parental RajiMex⁻ and

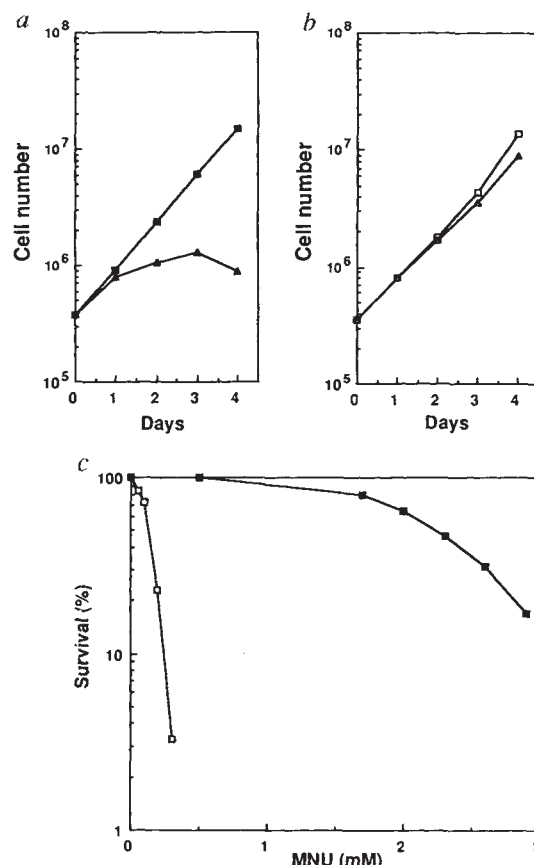


FIG. 1 Resistance of tolerant cells to MNU. The *m*⁶-G-DNA methyltransferase-deficient Burkitt's lymphoma line RajiMex⁻ has been described⁶. Several methylation-tolerant derivatives were isolated by multiple treatments with increasing concentrations of MNU. The surviving population that was resistant to MNU concentrations of up to 0.4 mM was cloned in 96-well plates. The survival of clone RajiF12 after treatment with MNU is shown. Cells were treated in suspension at 3×10^5 per ml in RPMI containing 10% fetal calf serum; MNU was not removed after treatment. Treated cells were seeded into 24-well plates, maintained in logarithmic growth by appropriate dilution and counted microscopically at daily intervals. a, RajiMex⁻ untreated (■) or treated with 10 μM MNU (▲); b, RajiF12 untreated (□) or treated with 0.3 mM MNU (△). The transfectant cell line CHOMT⁺ and its resistant derivative clone B have been described. Survival was determined according to ref. 8. c, CHOMT⁺ (□); clone B (■).

CHOMT⁺ cells. Extracts of the tolerant RajiF12 and clone B cells were selectively defective in G-T binding. An apparently nonspecific binding that is removed by an unlabelled matched competitor oligonucleotide, and binding to A-C or pyrimidine-pyrimidine mispairs are similar in extracts of both parental and tolerant cells (Fig. 2b; and data not shown) and serve as an internal control. Loss of MNU resistance by the clone B revertant BS11 is accompanied by the reappearance of G-T binding activity in extracts of BS11 cells (Fig. 2c), indicating that the MNU resistance is related to the absence of G-T binding.

† To whom correspondence should be addressed.

FIG. 2 Mismatch binding proteins in cell-free extracts. Preparation of cell-free extracts and binding to 20 fmol end-labelled 34-nucleotide oligomers of general sequence 5'-AGCTTGGCTGCAGGTNGACGGATCCCCGGAATT-3', where $N = G, A, T$ or C , annealed to complementary strands to create single-base mispairs was carried out as described¹⁰. Where indicated, cell extracts were preincubated for 5 min at 20 °C in mixtures containing, in addition to poly(dI-dC) (50 $\mu\text{g ml}^{-1}$), 40 fmol standard duplex 34-mer containing a G-C base pair at the mismatch position. This preincubation suppresses non-specific binding to the substrate oligonucleotide. Protein-oligonucleotide complexes were separated on 6% polyacrylamide gels. To improve the resolution of complexes, electrophoresis was continued until free oligonucleotide had migrated out of the gel. **a**, Binding to G-T mismatches by extracts of RajiMex⁻ and RajiF12: 15 μg extract was preincubated in the presence (lanes 2, 4, 6 and 8) or absence (lanes 1, 3, 5 and 7) of G-C oligonucleotide. End-labelled duplex oligonucleotide substrate containing a G-C base pair at the mismatch position (lanes 1, 2, 5 and 6) or a G-T mismatch (lanes 3, 4, 7 and 8) was then added and incubation continued for a further 20 min at 20 °C. Band C is an apparently nonspecific complex that is formed with all oligonucleotides and is removed by preincubation with cold G-C competitor oligonucleotide. Band B is mismatch-specific and is unaffected by the presence of a matched competitor. **b**, Binding to A-C mismatches. Preincubation and incubation were as in **a**, with the exception that the substrate oligonucleotides contained either an A-T (lanes 3 and 4) or A-C (lanes 1, 2, 5 and 6) base pair. Band A is the A-C- and pyrimidine-pyrimidine-specific mismatch complex. **c**, G-T binding by CHOMT⁺ and related extracts. Preincubation and incubation with extracts (15 μg) prepared from the parental CHOMT⁺ (lanes 1, 2), tolerant (clone B) (lanes 3, 4) and the revertant BS11 (lanes 5, 6) were carried out as in **a**. Lanes 1, 3 and 5: no G-C competitor. The migration positions of the nonspecific complex (band C) and the G-T (band B) complex are marked. **d**, Synergistic increase in G-T binding in mixed extracts. RajiMex⁻ extract was combined with different amounts of RajiF12 extract as shown. Binding to G-T oligonucleotide was then determined. All samples were preincubated with G-C competitor oligonucleotide. **e**, Absence of complementation of G-T binding by mixing mutant extracts. Extracts were combined as indicated and preincubated with G-C competitor oligonucleotide as described in **a**.



When RajiMex⁻ extracts were mixed with RajiF12 extracts, there was a synergistic increase in G-T binding (Fig. 2d), suggesting that complex formation might require two factors, one of which is limiting in RajiMex⁻ extracts, whereas the other is absent from RajiF12. RajiF12 and clone B are apparently defective in the same protein, however, because the mutant extracts are unable to complement each other and restore G-T binding activity when mixed (Fig. 2e). In a control experiment, mixing of RajiMex⁻ and clone B extracts also produced more-than-additive binding (data not shown), indicating that the components of the G-T binding complex from human and hamster cells are able to interact.

Mismatch correction-defective bacteria or yeast have high rates of spontaneous mutation^{4,11}. RajiF12 and clone B both show increased spontaneous mutation rates. The rate of mutation to hypoxanthine phosphoribosyltransferase deficiency (HPRT⁻) in RajiF12 was 3.6-fold higher than in RajiMex⁻, and mutation to adenosine phosphoribosyltransferase deficiency (APRT⁻) in clone B was 2.2-fold higher than the parental CHOMT⁺ line (Table 1). Loss of MNU resistance in the revertant BS11 was accompanied by a return to a mutation rate close to that of the parental CHOMT⁺ cells. The increases in spontaneous mutation rates in the tolerant lines are significant and reproducible; they

are moderate compared with the large increases seen in mismatch-defective bacteria⁴, although mutator mammalian cell lines generally have lower rates than mutator bacterial strains. In addition, deletions comprise many of the spontaneous mutations at these loci in mammalian cells^{12,13}. As mismatch correction protects predominantly against transition and single-base frameshift mutations^{14,15}, this bias in the mutational spectrum might attenuate the mutator effect in correction-defective mammalian cells.

These data provide direct evidence that mammalian G-T binding activity helps prevent spontaneous mutation and that tolerance to methylation damage is associated with a defect in this process. The precise role of the mammalian G-T binding activity is not known, but its ability to bind to mismatches is consistent with a role similar to that of MutS of *E. coli* and *Salmonella typhimurium* and to HexA of *Streptococcus pneumoniae*. MutS and HexA show extensive sequence homology to *Schizosaccharomyces pombe*¹⁶, human (Duc-1)¹⁷ and mouse (Rep-3)¹⁸ proteins, suggesting that function is highly conserved in this family of proteins.

We have observed two unrelated cell lines with impaired G-T binding associated with a mutator phenotype. In five more tolerant lines, mismatch binding by cell extracts was normal.

TABLE 1 Mutation rates in mismatch binding-defective strains

RajiMex ⁻ and RajiF12						Mutation rate (per cell per generation)
Cell line	Mutation frequency (×10 ⁻⁶)					
	Day 20	Day 36	Day 50			
RajiMex ⁻	4.5 ± 1.6	10.5 ± 5	16.5 ± 2.5		2.1 × 10 ⁻⁷	
RajiF12	23.0 ± 11.2	37.1 ± 9.5	42.7 ± 8.5		7.5 × 10 ⁻⁷	

CHOMT ⁺ and derived strains						Mutation rate (per cell per generation)
Cell line	Cultures	Number of resistant colonies per culture				
		Range	Mean	Variance	P ₀	
CHOMT ⁺	120	0-168	4.3	328	0.61	1.1 × 10 ⁻⁷ ± 0.2
Clone B (expt 1)	39	0-55	6.4	127	0.31	2.4 × 10 ⁻⁷ ± 0.5
Clone B (expt 2)	38	0-51	6.1	99	0.31	2.4 × 10 ⁻⁷ ± 0.5
BS11	40	0-225	12.8	43	0.50	1.4 × 10 ⁻⁷ ± 0.3

Three RajiMex⁻ and three RajiF12 subclones were isolated by plating in 96-well plates. Single colonies were picked and cultured. Exponential growth was maintained by daily dilution. On days 20, 36 and 50 after the initial subcloning, cells from each of the six subclones were plated at two different dilutions in duplicate 96-well plates in the presence of 5 $\mu\text{g ml}^{-1}$ 6-thioguanine. Positive wells were scored after a further 10 days incubation. Plating efficiency for all clones in non-selective conditions was determined similarly. Mutation frequencies were calculated from the Poisson distribution and were not significantly different at any of the sample times for each set of three subclones and data have therefore been pooled. Values are $\pm$ s.d. when appropriate. The generation time of each RajiMex⁻ and RajiF12 clone was determined from growth curves to be 18 h. Mutation rates in CHOMT⁺, clone B and BS11 were determined by Luria-Delbrück fluctuation analysis²⁰. Cells were plated in α -MEM containing 4 μM thymidine and 10% fetal calf serum at 100 cells per 6-cm plate and grown to 3×10^6 cells per plate. Each culture was then distributed equally among five 10-cm plates containing medium supplemented with 8-aza-adenine. The mutation rate (μ) was calculated from $\mu = MC^{-1} \ln 2$, where C is the number of cells plated in selective medium and M is the mean number of mutations per culture ($M = -\ln P_0$, where P_0 is the number of cultures containing no mutants). Standard deviations were calculated according to ref. 21.

Correction of mismatches is likely to involve the concerted action of several different proteins and these five lines could be defective in other steps in the same pathway. Elucidation of the functions of the *E. coli* mismatch repair proteins MutH, MutS and MutL was facilitated by defined mutant strains¹⁹. No simple selection for mammalian cell lines defective in mismatch correction exists but the isolation of methylation-tolerant cells is straightforward. The implication of defective mismatch correction in alkylation tolerance offers a new approach to investigation of DNA mismatch correction in mammalian cells. □

Received 24 September 1992; accepted 17 February 1993.

- Karran, P. & Bignami, M. *Nucleic Acids Res.* **20**, 2933-2940 (1992).
- Karran, P. & Marinus, M. G. *Nature* **296**, 868-869 (1982).
- Goldmacher, V. S., Cuzick, R. A. & Thilly, W. G. *J. Biol. Chem.* **261**, 12462-12471 (1986).
- Cox, E. C. *A. Rev. genet.* **10**, 135-156 (1976).
- Lieb, M. *Genetics* **128**, 23-27 (1991).
- Karran, P., Stephenson, C., Macpherson, P., Cairns-Smith, S. & Priestley, A. *Cancer Res.* **50**, 1532-1537 (1990).
- Aquilina, G. *et al. Carcinogenesis* **9**, 1217-1222 (1988).
- Aquilina, G., Zijno, A., Moscufo, N., Dogliotti, E. & Bignami, M. *Carcinogenesis* **10**, 1219-1223 (1989).
- Jirincy, J., Hughes, M., Corman, N. & Rudkin, B. B. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8860-8864 (1988).
- Stephenson, C. & Karran, P. *J. Biol. Chem.* **264**, 21177-21182 (1989).
- Borts, R. H. *et al. Genetics* **124**, 573-584 (1990).
- Phear, G., Armstrong, W. & Meuth, M. *J. molec. Biol.* **209**, 577-582 (1989).
- McGregor, W. G., Maher, V. M. & McCormick, J. J. *Somatic Cell and Molecular Genetics* **17**, 463-469 (1991).
- Rewinski, C. & Marinus, M. G. *Nucleic Acids Res.* **15**, 8205-8215 (1987).
- Schaeper, R. M. & Dunn, R. L. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6220-6224 (1987).
- Fleck, O., Michael, H. & Heim, L. *Nucleic Acids Res.* **20**, 2271-2278 (1992).
- Fujii, H. & Shimada, T. *J. Biol. Chem.* **264**, 10057-10064 (1989).
- Linton, J. P. *et al. Molec. cell. Biol.* **9**, 3058-3072 (1989).
- Modrich, P. *J. Biol. Chem.* **264**, 6597-6600 (1989).
- Luria, S. E. & Delbrück, M. *Genetics* **28**, 491-511 (1943).
- Lea, D. E. & Coulson, C. A. *J. Genet.* **49**, 264-285 (1949).

ACKNOWLEDGEMENTS. We thank C. Stephenson for technical assistance in the early stages of this project. This work was supported in part by a CNR ACRO award and a European Community grant to M.B., a Royal Society/CNR joint travel grant, and by the Imperial Cancer Research Fund.

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without interruptions, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. High priority cannot be given to pre-submission enquiries; in urgent cases they can be made in the form of a one-page fax. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are typeset from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose, and to mention availability of these data.

Once a manuscript is accepted for publication, contributors will receive proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check proofs carefully. Manuscripts are generally published 2-3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50-80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few short subheadings.

Letters are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and should contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned but suggestions can be made to the Commentary Editor in the form of a one-page synopsis.

News and Views articles inform non-specialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views Editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and four copies are required, each accompanied by artwork. If photographs are included, five sets of originals are required; for line drawings, one set of originals and four good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Three copies of relevant manuscripts in press or submitted for publication elsewhere should be included with each copy of submitted manuscripts, and clearly marked as such. Five copies of revised and resubmitted manuscripts, labelled with their manuscript numbers are required, together with five copies of a letter detailing the changes made.

Titles are brief and simple. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963-65). The first and last page numbers are included; reference to books should include publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used in the text.

Acknowledgements are brief and appear after the reference list; grant and contribution numbers are not allowed.

Supplementary information is material relevant to Articles or Letters which cannot, for lack of space, be published in full, but which is available from *Nature* on request.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1234 National Press Building, Washington, DC 20045, USA. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax.

Unfair cops?

Steve Blinkhorn

The Psychology of Interrogations, Confessions and Testimony. By Gisli Gudjonsson. Wiley: 1992. Pp. 362. £24.95, \$48.95.

BUDDING Torquemadas stop here and look elsewhere for a thrill. This is the story of everyday inquisitorial folk, seeking conviction more by confession than by detection and denied the aid of rack and thumbscrew. It is also, coincidentally, an object lesson in how a soft, social, semi- and sometimes pseudo-science (for such is psychology) can carry the day when set against poorly controlled real science.

Gisli Gudjonsson is an eminent forensic psychologist whose original career was as a policeman in Iceland. He has written a book that is a model of academic rectitude in crediting original sources, judiciously weighing the relative merits of competing theories and maintaining a detached and dispassionate style when discussing his own work. On the surface it is a worthy academic tome. But its subject matter is life and death, and decades of imprisonment, and the perpetuation of grudge, and the philosophy of human thought and action under the extreme pressure of interrogation under suspicion of serious crime.

People confess to crimes they have not committed — or at the very least, the police have claimed that people have confessed to crimes they have not committed. Why? It will not do to say that, in the absence of sufficient evidence, the first recourse of the police is to invent some more. The police have ways of making people talk, and perhaps of getting them to say what they (the police) would like to hear. Those interested in finding out how to do this should refer immediately to Chapter 3.

When people confess to crimes they have not committed, psychologists disagree as to why they do so, even when they agree as to which confessions are false. The courts, meanwhile, are faced with varieties of expert testimony that they are ill-equipped to choose between and where the credibility of the expert may weigh more heavily than the credibility of the testimony or the adequacy of the science on which it purports to be based.

Attempts to recruit science into differentiating truth from falsehood have had at best a mixed history. Various

forms of truth serum, polygraph lie detection and statistical and linguistic analysis of written documents have all been pressed into service with the aim of improving the quality of decisions in criminal justice systems — or at least of increasing the clear-up rate of crimes, which may or may not be the same thing. The past year has seen the execution of a convicted murderer in the United States a matter of hours after submitting himself to a polygraph lie detection test (and 'failing'), and the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Facing the truth — 1937 Italian picture of a lie detector.

acquittal on appeal of Judith Ward in the United Kingdom, whose conviction of a terrorist atrocity nearly two decades ago was based largely on a confession, supported by now-discredited forensic evidence, to point to just two of the distressingly long list of wrongful and doubtful convictions.

It is hard to escape the conclusion that too much trust has been placed in forensic evidence based on methods derived from the physical sciences. Tests for the presence of traces of explosives on suspects' hands which prove sensitive to traces of inoffensive everyday substances are as deserving of disrepute as the polygraph, whose capacity for the detection of deceit is exceeded only by its capacity for falsely implicating the innocent. Enthusiastic science so often becomes bad science, and any kind of science sits uneasily in a courtroom.

Both the judicial process and the work of the scientist are centrally concerned with the evaluation of evidence and standards of proof, but in the end a court is a common-sense forum so far as the evidence is concerned — the law may be a different matter. It's not *what* you believe, it's *who* you believe.

What is ultimately less than satisfying about Gudjonsson's book is not the quality of his account but the state of the subject matter. We have come a long way in our understanding of language since Wittgenstein's *Tractatus*, but common sense, and by extension the police and the courts, may still assume there is a unique and natural point-to-point correspondence between events and reports of events. What psychologists have to say about the fallibility of memory and the suggestibility of people — even though this is not yet a single settled body of knowledge — tends to relegate the notions of truth and falsehood, like sanity, to the category of purely legal concepts, but does not entirely succeed. It is not a satisfying substitute for what we used to believe. The attraction of the lie detector is that it seems to restore the moral certainty that twentieth-century philosophy and psychology have stripped from our judgements.

There remains the question of the status of the expert witness, called to testify to the credibility of statements made under interrogation. It is hard to shake off the sense that, at least some of the time, psychologists base testimony on theoretical positions that will soon seem laughable. So one ends up preferring provisional theorizing from a credible source, but with an uncertain shelf-life, to hard science badly done.

Perhaps we all hanker after the days when the gifted amateur detective would confront the suspects in the library and reveal the culprit with an ineluctable line of reasoning and a single incontrovertible piece of evidence. Popular detective fiction from Conan Doyle on has portrayed its heroes as quasi-scientists who so often eventually extract a confession after an elegant process of deduction from evidence. The truth always was messier than fiction, and Gudjonsson reveals the messy truth about interrogation and confession: his welcome and timely book should be required reading for all who strike public attitudes on crime and punishment. □

Steve Blinkhorn is at Psychometric Research and Development Ltd, Brewmaster House, The Maltings, St Albans, Hertfordshire AL1 3HT, UK.

God's gaze

William R. Shea

Catching the Light: The Entwined History of Light and Mind. By Arthur Zajonc. Bantam: 1993. Pp. 388. \$24.95, £16.99.

WHAT is the nature of this invisible thing called light whose presence calls everything into view — except itself? Arthur Zajonc offers not a definitive answer but a fascinating account of what the question has meant over the ages. For the Ancient Egyptians, who sought a moral or spiritual answer, light was the gaze of God. For thinkers in the seventeenth century, who looked for a mechanistic account, it was explained by studying the geometry of refraction. In both cases, the quest was one-sided and, if this book has a moral, it is that we should read the full biography of light and not only the fragments that our own scientific community has written. Sight entails much more than just opening our eyes.

A few examples of colour vision among the Ancient Greeks and Romans convey some idea of the broad cultural and psychological basis of Zajonc's interdisciplinary study. Euripides writes in *Iphigenia* that blood made the side of the altar *ksanthon*. The word would therefore seem to mean red and Latin translators thus rendered it *ruber*. But Aeschylus in *The Persians* calls the leaves of the olive tree *ksanthon*, and in the *Iliad* the term is used for honey. Other colours receive an equally perplexing treatment at the hands of the Greeks. Both the Moon and fiery eyes are *glaukon*, so it is natural to think of this word as referring to yellow or red. But Plato uses it for the mixture of dark blue and white, associating it with sky-blue or blue-grey — *caesius*, as the Latins rendered it — whereas Euripides uses it for a vestibule decorated with leaves, relating it to the colour of vegetation. Translators seek a way out by rendering *glaukon* as 'brilliant' or 'shiny'. For instance, "glaukopis Athene" becomes "flashing-eyed Athene" in the Loeb Classical Library. But this solution is itself highly problematic because the eye disease we call 'glaucoma' does not give the eye a shiny but a turbid appearance. Perhaps the Greeks were colour-blind. But this radical solution is a counsel of despair. What is needed is not an eye doctor but a cultural anthropologist.

The Romans are closer to us and cannot be suspected of collective colour blindness. Those assuming that the English for *purpureus* is purple would find themselves in linguistic trouble when they discover that the poet Horace calls a swan *purpureus*. Indeed, the

Latins gave shiny black as well as shiny white objects the name purple. Or consider *caeruleus*, the word used to describe the sky, the sea and the eyes of Germans as well as the hair of Indians, the night and death! It is clear from these illustrations that a variety of different conceptual frameworks can be superimposed with equal validity on the order of 'coloured things'. The 'colours' that we encounter do not by themselves determine how we are to think of them independently of the apparatus of thought that we bring to the transaction.

Zajonc's catholic tastes and his rare gift of vivid characterization enable him to resurrect discarded images of vision such as the belief (shared by Plato, Euclid and even Galileo) that a visual ray emanating from the eye is essential for sight, or Aristotle's view that nothing streams from either eye or object but that illumination is a state or quality of the medium. Goethe's controversial ideas on colours are also displayed in their best light, and Rudolf Steiner's mystical musings on vision are examined against the interesting background of his quest for a science that would not belie genuine spiritual experience. The fundamental contributions of Faraday and Maxwell are lucidly described, and cur-

rent research in quantum optics is given a provocative slant. We are reminded throughout that theories are aids to reflection and should never be hardened into permanent and disabling truths.

Catching the Light is a treasure-house of telling quotations and it is a pity that the generally useful references are sometimes incomplete. In a work that ranges so far and wide, it is understandable that a few mistakes should have slipped in. The Berlin suburb of Grunewald, which is mentioned, indeed celebrated, at least a couple of times, is provided with a gratuitous umlaut. Descartes did not write his *Discourse on Method* in 1619 but in 1636. It is misleading to suggest that he explained the rainbow in terms of white and black, and although he could have written "Give me motion and extension and I will construct the universe", it seems more *ben trovato* than *ben documentato*. These and other trivial errors could easily be corrected to make the second edition not only a vastly enjoyable and stimulating work but a truly reliable one as well. □

William R. Shea is in the Centre for Medicine, Ethics and Law, McGill University, 3690 Peel Street, Montreal, Quebec H3A 1W9, Canada.

An essayist worth his saltations

Bryan C. Clarke

Eight Little Piggies: Reflections in Natural History. By Stephen Jay Gould. Norton/Jonathan Cape: 1993. Pp. 507. \$22.95, £18.99.

OH for J. B. S. Haldane's originality, Arthur Cain's encyclopaedic knowledge and Peter Medawar's skill in writing. Failing those, I would settle for Stephen Jay Gould's array of talents, perhaps with an added *souçon* of modesty and rigour. As the old proverb says, "If wishes were butter-cakes, beggars might bite". But you do not need to be a millionaire to appreciate a good meal, and this is a feast.

When Gould began writing, his essays were spottily pretentious, showing too obviously that he was a very clever fellow, sometimes with humorous results when they misfired. Moreover, it was difficult to forgive him, in his early advocacy of 'punctuated equilibria', for doing less than justice to the ideas of Simpson and Mayr. Even then, though,

he was patently struggling to perfect his art, and he consistently improved.

Time has passed, and the ugly duckling has grown into a swan, still with some neotenic imperfections it is true, but a veritable swan. His latest collection is a lovely mixture of bizarre facts, nice arguments, clever insights into the workings of evolution and a quality of writing that can make your skin prickle. Surely enough there is a trick to it. You start with a small commonplace or oddity, weave it into an exemplar of some general principle, patch in a few quirky or informative asides and then come back to view the beginning from a new angle. It is a trick, but a difficult trick to do, and Gould does it supremely well.

Inevitably he is uneven. My own favourites are the chapters on vertebrates' jaws, ichthyosaurs' tails and the head-shields of cephalaspis fishes. He seems less secure in three essays based on the fossil fauna of the Cambrian period. The first, about meeting someone who knew Charles Walcott (the discoverer of the Burgess Shales), leaves an unpleasant taste because it seems to patronize a modest and respected scientist. The second and third concern the animals themselves. We learn that two puzzling 'shelly' fossils, *Halkieria* and

■ Cambridge University Press has recently published volume 8 of *The Correspondence of Charles Darwin* edited by F. Burkhardt, D. M. Porter, J. Browne and M. Richmond. The volume covers 1860, a year dominated by the public and private response to the *Origin of Species*. £40, \$59.95.

Microdictyon, are respectively spicules from organisms that look suspiciously like amphineuran molluscs (although he does not say so) and plates from the sides of onychophorans.

The strangest of the Cambrian wonders, *Hallucigenia*, turns out to be an onychophoran too, not immediately recognized because it was interpreted upside down. In his earlier book, *Wonderful Life*, Gould argued that the Cambrian fauna, with extraordinary creatures such as *Hallucigenia*, displayed a level of anatomical diversity unequalled thereafter, and that the persistence of modern phyla was a matter of historical 'contingency'. Now, he uses the relegation of *Hallucigenia* to a modern phylum in making a case for the same view. One cannot avoid discomfort with a theory that seamlessly accommodates so complete a reversal of fact. Elsewhere there are other weaknesses. He does not realize that duplicating a gene can be favoured by natural selection either through a need for more of its product or as an escape from heterozygous advantage. He falls into the common trap of supposing that only the neutral theory predicts a faster evolutionary rate for smaller changes of phenotype, whereas selective theories do so too.

Gould is very good at finding sense in the work of scientists whose ideas have now been abandoned and who are often ridiculed in textbooks. The pieces on William Paley and Archbishop James Ussher stand out, but he is palpably ill at ease trying to justify Goethe's amateurish theories about the structure of plants. There are also some fine essays on general subjects, including a heartfelt and perceptive celebration of authenticity. Having read the piece, and mentally applauded, I chanced to look more carefully at the dustjacket. Here, in the US edition, there is a picture of some snails and the caption: "These sixteen Partula snails from Moorea . . . have all become extinct in the last ten years". The shells illustrated are *Partula nodosa* from Tahiti. So much for authenticity. Another piece tells how Gould was dismayed to find that he had misremembered the place of treasured meetings with his grandfather. It is true and touching, but in an earlier collection of essays he accused Teilhard de Chardin of the Piltown forgery because of a lapse no greater than his own. These peccadillos, however, are forgivable in one who writes so well. Gould has given us a feast, and the reader should enjoy it, taking some of the arguments with a pinch of salt. After all, a little salt improves the savour. □

Bryan C. Clarke is in the Department of Genetics, University of Nottingham, Nottingham NG7 2UH, UK.

Heavenly works

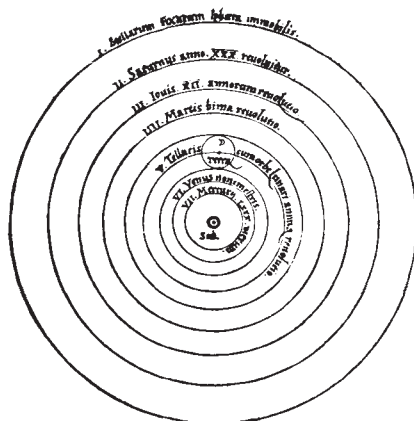
Fred Hoyle

Nicholas Copernicus: Complete Works in Two Volumes*. Johns Hopkins University Press: 1993. Pp. 452/373. Each volume \$48, £30 (pbk).

Johannes Kepler: New Astronomy†. Cambridge University Press: 1992. Pp. 665. £85, \$140.

BOTH publishers are to be congratulated on making these works, which form the foundations of modern astronomy, available to English readers uneasy with mediaeval Latin. Those who venture

NICOLAI COPERNICI
net, in quo terram cum orbe lunari tanquam epicyclo contineri
diximus. Quinto loco Venus nono mente reduciatur, Sextum
denique locum Mercurius tenet, octuaginta dierum spacio circū
currentis. In medio uero omnium refidet Sol. Quis enim in hoc



palcherimo templo lampadem hanc in alio uel meliori loco ponere, quam unde totum simul posuit illuminare: Siquidem non inextinguibilem lucernam mundi, alij mentem, alij rectorem uocant. Trimegitus utilis Deum, Sophodis Electra inuenit omnia, la profecto tanquam in folio re gali Sol refidens circum agens gubernat Astrorum familiam. T ellus quoque minime fraudatur lunari ministerio, sed ut Aristoteles de animalibus ait, maxima Luna cum terra cognatione habet. Concipit interea à Sole terra, & impregnatur annuo partu. Inuenimus igitur sub hac

The first printed representation of the Copernican system as it appears in *De revolutionibus* (1543). Picture taken from *Early Physics and Astronomy* (2nd edn) by O. Pedersen. CUP, £50, \$75 (hbk), £19.95, \$27.95 (pbk).

between the covers of any of the volumes will be surprised to find out how much more complicated the work looks than popular potted versions would have it. Compilers of the popular versions, I suspect, do not really understand what the original authors sought to explain. But seen as the pioneers had to see it, the situation was devilishly complicated, as much so as any of the puzzles of recent science.

Two versions of *On the Revolutions* exist, one made in Copernicus's last

years by Rheticus, a visitor from Germany whose copy was used for the printed book set in Nürnberg in 1543; the other Copernicus's own autograph. Both versions were used to obtain what the translator, Edward Rosen, believes to be a superior rendering than either alone. Frankly, I would have preferred Copernicus's own manuscript, since it is not clear that the Rheticus copy ever had Copernicus's full approval.

It is scarcely possible to consider the constructions of Copernicus without comparing them with those of Ptolemy, who seems to me to have been dealt with more than harshly by scientific history. Copernicus discovered a geometrical construction for obtaining the motion of a planet, correct in the main to the first order in the eccentricity of its orbit, working in heliocentric coordinates. Then, to obtain a planet's position with respect to the Earth, in order to compare theory and observation, a subtraction had to be made, the planet's position minus the Earth's position. Ptolemy, on the other hand, discovered a construction that both gave the planet's position and made the subtraction with respect to the Earth, all in the same process. The outcome, so far as comparing theory and observation was concerned, was essentially the same. The science was the same. It was the inconsequential philosophy that people attached to the theory that was different.

Two small mistakes in the constructions of Copernicus provided the springboard for Kepler. In his construction for planets other than the Earth, Copernicus used a point related to the Earth when he should have used the Sun, a strangely non-copernican oversight. Kepler, when he had withstood the insults of Tycho Brahe and had eventually secured access to Brahe's observations, soon put the mistakes right, and might have rested content there. But discrepancies remained for Mars. Because the eccentricity of Mars is comparatively large, errors due to second-order terms were showing through in Brahe's data. How Kepler solved this problem is one of the great early stories of science, on a par with the "Eureka!" of Archimedes. Just what the solution was remains unexplained in Dreyer's classic *History of Astronomy*. Nor does it appear in Kaspar's biography of Kepler. And the avid reader will have a high old time digging it out of *New Astronomy*. Some hints may help.

Kepler's first move was to use an opposition of Mars (a situation when the terrestrial observer knows the heliocentric direction of Mars with respect to the distant stellar background) together with routine observations of Mars, to obtain the relative shape of the Earth's orbit to the highest possible accuracy. Then other oppositions of Mars, making use of

* *On the Revolutions* edited by J. Dobrzycki, translation and commentary by E. Rosen; *Minor Works* edited by P. Czartoryski, translation and commentary by E. Rosen with E. Hilfstein. † Translated by William H. Donahue.

the accurately known sidereal period of Mars, gave heliocentric directions and distances for a number of points on the martian orbit. And the last step, the one that commentators tend to feel more comfortable with, was to decide the nature of the curve determined by the points. The usual story is that Kepler found the curve to be an ellipse, to which the mention of 'perfect ellipse' in the title of Chapter 59 of *New Astronomy* gives credence. But with a proper measure of respect, this was an exaggeration. The martian orbit could be said to be an ellipse only to within the accuracy of Brahe's observations, and these were far from perfect.

Underlying every investigation by all the pioneers was the hypothesis that the planetary orbits (neglecting mutual interactions) were closed curves, requiring a unique relation between direction and distance as viewed by the Sun. It was

just lucky that the pioneers did not have to cope with a world in which the post-newtonian relativistic approximation was important, when this hypothesis would have been untrue.

Quite possibly, Newton and Halley were aware of this point. Then one can readily understand the excitement over the orbits of comets of large eccentricity, which really did require the close perfection of the elliptic orbit. This matter of deep principle is surely the reason why comet Halley was so important to the newtonian world. For the perfect ellipse implied the inverse-square law of gravitation, and with that relationship firmly established, the march of science had become essentially unstoppable. It was in these books that it all began. □

Sir Fred Hoyle is at 102 Admiral's Walk, West Cliff Road, Bournemouth, Dorset BH2 5HF, UK.

The shape of things to come

John L. Casti

Predicting the Future. Edited by Leo Howe and Alan Wain. *Cambridge University Press: 1993. Pp. 195. £18.95, \$29.95.*

WITHOUT too great an exaggeration, I think it's fair to say that the *raison d'être* of intellectual endeavour in general — art, literature, religion, music and all the rest — is to offer convincing answers to the eternal question: "Why do we see what we do and not see something else?" And science is no exception, being distinguished only by the form of the answers the scientist provides. The scientific answers come wrapped up in a set of rules, usually encapsulated in the form of a mathematical model or, more common nowadays, a computer program. And the purpose of these rules? Basically, to do two things: to explain past observations and to predict the results of new ones. So prediction and explanation are the twin pillars upon which the scientific enterprise rests. But, as this book makes clear, prediction and explanation are not concerns of the scientist alone, but lie at the centre of all types of intellectual life ranging from astrophysics to the Holy Scriptures.

Predicting the Future arose out of a series of lectures on the topic of "Predictions" given at Darwin College, Cambridge, in 1991. But from a glance at the table of contents, the volume might have been better titled "Predictions, Prophecies and Futurology". The book opens well enough, with thoughtful and informative accounts of the long-term be-

haviour of the Universe by Stephen Hawking and the mysteries of chaotic dynamical processes by Ian Stewart. Both of these well-written essays are firmly centred in the scientific tradition of prediction by rule, showing that the rules we currently have at hand leave a lot to be desired when it comes to addressing the questions we most want to ask of nature.

Following a historical essay on the role of comets in prophecy through the ages by Simon Schaffer, the distinguished economist Frank Hahn offers probably the best account of the ins and outs of prediction in the book with his chapter on divining the mysteries of economic processes. Along the way to his almost foreordained conclusion that the predictive powers of economists are pretty feeble, Hahn brings out an important although often neglected aspect of prediction: namely, that not all predictions involve the future. He supports this claim with several convincing examples from the area of rational choice theory. Hahn ends with a general discussion of the possibility of effective predictions in the social sciences in general, a discussion that should sober up just about anyone infected with the germ of the idea that there's anything even remotely scientific about the social 'sciences'.

At this point, the book veers off the path of anything even faintly resembling prediction — scientific-style — flying off into the realms of futurology, historicism and prophecy. The first entry in this direction is an account of the possible future of medicine by Ian Kennedy. In pondering the medical frontier, Kennedy

focuses on the bioethical dilemmas that the medical community is likely to face in coming years, including such things as the complex questions of care for an ageing population, how to handle developments in human genetics and the control of access to information about patients and their medical care.

The final three chapters abandon completely any pretence at prediction as that term is used in everyday speech, giving the floor over to religion, science's main competitor in the reality-generation game. Averil Cameron fires the opening salvo with his account of divine providence in late antiquity. This turns out to be a not very convincing attempt to show how late antiquity was characterized by a hierarchical Christian worldview that claimed to offer answers to a wide spectrum of human problems and was an infallible guide to predicting the future. The next chapter by Richard Gombrich on the role of prediction in the Buddhist scheme of things is a bit more satisfying, mostly because the author restricts his attention to a simple question: how open is the future from a Buddhist's perspective? Gombrich offers a threefold answer, in which the future can be both open and closed depending on whether or not an individual is one of the 'enlightened'. Lastly, there is the Last Judgement, the theme of the book's final chapter by Don Cupitt. In this strange chapter, the author traces the history of the almost universal belief that there is a moral providence in the world ensuring that we will eventually get exactly what we deserve.

So what is one to make of this potpourri? To begin with, it's difficult to think of any single individual short of a reincarnation of Leonardo da Vinci whose intellect could simultaneously encompass the wide variety of views espoused here on the nature and effectiveness of prediction. Most scientists will find the second half of the book distinctly odd, if not completely eccentric, in treating prediction as more of a kind of religious and/or historical experience than as the working out of a set of scientific rules, or a computer program, to get at the why and the when of things. All in all, this seems to be a volume more for browsing than for line-by-line reading. While the editors are to be commended for bringing such a disparate collection of views together within a single volume, it remains a volume in which the whole is somehow less than the sum of the parts. □

John L. Casti is in the Institute for Econometrics, Operations Research and Systems Theory, Technical University of Vienna, Vienna, Austria, and the Santa Fe Institute, Santa Fe, New Mexico 87501, USA.

Absinthe makes the art grow stronger

John Mollon

Vincent van Gogh: Chemicals, Crisis, and Creativity. By Wilfred Niels Arnold. Birkhäuser: 1992. Pp. 332. DM98, £37.20, \$49.50.

BOOKS on the ailments of the distinguished dead form a recognizable literary genre; and there is a whole sub-genre devoted to the madness of Vincent van Gogh. During the last three years of his life — a period that saw the painting of some of his finest works — van Gogh had several episodes of mental disturbance. Among the diagnoses offered have been schizophrenia, idiopathic epilepsy, manic-depressive psychosis, syphilis, Ménière's disease and intoxication by metal salts. Psychoanalysts still trace his instability to his relationship with his parents and to the stillborn infant, Vincent, who preceded him by one year. Wilfred Arnold's hypothesis is that van Gogh suffered from the rare metabolic disorder acute intermittent porphyria, and that his attacks were precipitated by indulgence in absinthe.

Acute intermittent porphyria arises from mutations in the gene that codes for porphobilinogen deaminase, the third enzyme in the pathway for the biosynthesis of haem. The inheritance is autosomal dominant with variable penetrance. An almost invariable symptom is abdominal pain, and muscular weakness, psychosis and seizures are common. The attacks begin after puberty and are intermittent; and indeed around 90 per cent of heterozygotes may pass through life with only half the normal level of the enzyme and yet not show symptoms. Under unstressed conditions, it is thought, the rate of flux of substrates through the haem pathway is below the threshold at which the affected enzyme becomes rate-limiting. Any increased demand for haem in the liver will induce δ -aminolevulinic acid synthase, make porphobilinogen deaminase rate-limiting and lead to the accumulation of intermediates. Fasting or drugs may precipitate such an attack.

Van Gogh's doctors have left us only the tersest clinical notes, but the artist was himself much concerned with his health and it is a frequent topic in his letters to his brother Theo. Arnold supports his theory with a detailed examination of the letters; and, as his theory requires, he also assesses the evidence that Theo and other members of the family suffered from the same hereditary condition. He supposes that van Gogh's attacks were precipitated by neglect of meals, by the use of camphor against insomnia and in particular by drinking absinthe, the fashionable clear green

liqueur that turned to an opalescent pale yellow when diluted with water. An absinthe bottle appears in more than one of van Gogh's paintings, and we have Paul Signac's account of how van Gogh "would take his seat on the terrace of a café. And the absinthes and brandies would follow each other in quick succession." In the southern department of Bouches-du-Rhône, where van Gogh lived from 1888 to 1891, the average annual consumption of absinthe was a

(George III and the Mad-Business, Allen Lane, 1969; curiously uncited by Arnold) were able to find contemporary reports of discoloured urine not only in the case of George III, but even in the case of his seventeenth-century ancestor James I. Arnold has found no such reference in van Gogh's letters. He explains this away rather lamely: van Gogh probably pissed in the fields and his urine seldom had the chance to stand around and turn the opaque port-wine colour that reveals the presence of porphyrins. Yet van Gogh lived in the heyday of the *vase de nuit* and his crises were largely spent in the hospital at Arles or the asylum at St Rémy.



Damaging diet — van Gogh's *Still Life, Drawing Board with Onions, Raspall's Book, Absinthe Bottle etc.* (1889, St Rémy, oil on canvas).

remarkable 2.45 litres per head, and it was only in the twentieth century that the toxic effects of the liqueur led to a ban. Its most sinister ingredient was thujone, a terpene derived from wormwood (*Artemisia absinthium*). A standard test for drugs that induce porphyria is to measure the induction of δ -aminolevulinic acid synthase in liver cells from fetal chickens; and Arnold reports a positive result when testing thujone and camphor.

It makes a good read. But ultimately the evidence is circumstantial. Although Arnold lists a number of letters that refer to 'gastrointestinal complaints', the relevant passages in fact speak mostly of 'weakness of the stomach' rather than severe abdominal pain. And notably missing from this detective story is the smoking urine. I. MacAlpine and R. Hunter in their study of putative porphyrics in the British royal family

The polymerase chain reaction will take some of the sport out of this genre of medical history. For the famous and the infamous have often left fragments of themselves behind, and it is possible nowadays to amplify sections of suspect genes from only micrograms of DNA — allowing a test, for example, of whether Abraham Lincoln owed his tallness to Marfan's syndrome (see V. A. McKusick, *Nature* 352, 279; 1991). The gene for porphobilinogen deaminase has been sequenced and several mutations that give rise to acute intermittent porphyria are already known. So that amputated ear, had it been preserved, would have given us a firm answer. But the hospital at Arles kept it for a year and then threw it away. □

John Mollon is in the Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK.

Fire and brimstone

Jonathan Fink

The Lyman Hawaiian Earthquake Diary, 1833–1917. By M. Wyss, R. Y. Koyanagi and D. C. Cox. *US Geological Survey Bulletin 2027, USGS: 1992. Pp. 34. \$2.75 (pbk).*

Hawai'i Volcano Watch: A Pictorial History, 1779–1991. By T. L. Wright, T. J. Takahashi and J. D. Griggs. *University of Hawaii Press: 1992. Pp. 162. \$42 (hbk), \$24.95 (pbk).*

FEW things instil the fear of God like a volcanic eruption. That may be why missionaries flock to places where volcanoes dominate the landscape, such as Latin America and the Pacific Islands. Although many of these religious salespeople rely on the fear of natural catastrophes to supply them with customers, others strive to understand their natural surroundings. So it was with early Christian settlers in the Hawaiian islands, several of whom used soul-saving as an excuse to explore the local volcanoes. Two recent publications illustrate how accurately their observations presaged those of modern scientists, and even provided a unique dataset still useful today for predicting future eruptive activity.

Sarah Lyman, a missionary wife transplanted from Vermont, would seem to be an unlikely candidate for chief Hawaiian seismologist. Yet a diary she and her daughter-in-law kept between 1833 and 1917 now represents the only continuous record of Hawaiian earthquake activity in the nineteenth century. The diary has recently been published with a recalibration and interpretation as *The Lyman Hawaiian Earthquake Diary, 1833–1917*. This valuable compilation allows the modern Hawaiian seismic re-



Kilauea by moonlight, painting by E. Bailey (1901).

cord to be extended back to before the giant Kau earthquake of 1868, the largest ever documented for the islands.

Wyss and his colleagues convert the Lymans' descriptive entries (for example, "1875 Jan 29th: A smart shake at 12:15 M. preceeded some 12 seconds by a sudden bump & followed by several seconds of trembling") to the Modified Mercalli Intensity Scale, which relates earthquake shaking to its effect on man-made structures. They then plot the way in which the number of intense Hawaiian earthquakes has varied from 1832 to the present. The data reveal a relatively quiet period before the major 1868 earthquake, a very active interval from 1868 to the start of modern record-keeping in 1915, and an intermediate level after that. Clearly, the 1868 event caused changes in the volcano's structure that were felt for nearly five decades.

Recognizing such long-term trends is an essential part of any volcanic forecast because they help define cyclic patterns not evident in short-term records.

The Lymans were part of an inquisitive group of nineteenth century Hawaiian missionaries, explorers and scientists who were lured to eruptions by the same fascination that draws modern tourists, journalists and geologists. *Hawaii Volcano Watch* is a delightful new book that incorporates the impressions of these early observers into a vividly illus-

trated chronicle of Hawaiian volcanology, from the early Polynesians' worship of the volcano goddess Pele to contemporary monitoring using lasers, satellites and supercomputers.

Historical perspective, enhanced by an outstanding collection of paintings, sketches, maps and photographs, two examples of which are shown here, offers lessons for scientists and Hawaiian residents alike. Geologists may become less smug on learning that many of their most important 'recent' findings, such as the age progression of the island chain or the role of lava tubes in producing very long flows, were actually proposed more than a century ago. Retired people from the mainland drawn to vacation homes on Kilauea's slopes may hesitate when reminded that more than 90 per cent of the volcano's land surface has been covered by lava since the arrival of the first Hawaiians some 1,500 years ago. And developers building palatial resorts along the beaches of Hawaii's Kona coast and the island of Maui, out of view of Kilauea's frequent outbursts, won't want to publicize the fact that lava flows have threatened these areas within the geologically recent past 200 years.

Geology is based on the principle that the present is the key to understanding the past. By highlighting the prescient contributions of early volcano watchers, these two books turn the maxim around: for Hawaiian volcanology, the past has often been the key to the present. □

Jonathan Fink is in the Division of Earth Sciences, National Science Foundation, 1800 G Street NW, Washington DC 20550, USA.



Lava from a Kilauea eruption covers the beach east of Kupapau Point. (USGS photo by J. D. Griggs, 1988.)

The gadfly

Gregg Herken

Genius in the Shadows: A Biography of Leo Szilard — The Man Behind the Bomb. By William Lanouette with Bela Silard. Scribner's: 1993. Pp. 587. \$35.

LEO Szilard was one of a remarkable group of Hungarian physicists who emigrated to the United States in the 1930s and played a pivotal role in the making of the atomic bomb. "If the uranium project could have been run on ideas alone, no one but Leo Szilard would have been needed", observed Eugene Wigner, Szilard's countryman and colleague. A difficult child, Szilard grew up to become an impossible adult; one whose preferred activity was creative daydreaming — "botching", he called it — during long baths and in more unlikely settings. It was on one such occasion, while crossing a London street in 1933, that he saw how an atomic chain reaction might happen. He filed the first patent for an atomic reactor with the British Admiralty several months later.

Characteristically, however, Szilard himself refused to do the experiments necessary to discover which element or elements could undergo fission, preferring, instead, to importune colleagues into carrying out the work. His lassitude had not prevented him from escaping from Germany just one day before the Nazis closed the border, but it did keep him from being the first to find out that uranium was the element he was after. Years later, he joked that he and his fellow procrastinators should have received the Nobel prize — for peace, not physics.

Readers of William Lanouette's sympathetic but not hagiographic account of Szilard's life, researched with the assistance of Leo's brother, may find themselves wondering why the author is so obviously fond of his subject. "I've always been a great man", Szilard once sneered to a would-be admirer who was trying to pay him a compliment.

In truth, Szilard's greatness stemmed from the fact that he was a catalyst rather than a prime mover; a facilitator who acted as an academic matchmaker at critical times to bring competing ideas, or scientists, together. Most famously, in the summer of 1939, he persuaded Albert Einstein to sign a letter, which he himself had drafted, warning President Franklin Roosevelt about the danger of atomic energy. Successful in this endeavour, Szilard imagined himself becoming the high priest of a scientific association, or *Bund*, that would not only build the bomb but advise, and perhaps even decide, on its use.

The fate of Szilard's atomic *Bund* was emblematic of the problem with his approach. Having put the process in motion, he sat back to criticize those doing the work. Unfortunately for him, this tendency towards high-handedness — and his self-professed delight in "baiting brass hats" — cast his subsequent career, and influence, permanently into the shadows. So irked was the director of the Manhattan Project, General Leslie Groves, with Szilard's interference, that Groves tried to have him interned for the duration of the Second World War. Failing that, Groves had Szilard



Leo Szilard (1898–1964).

tailed by a phalanx of counter-intelligence agents, whom the physicist took a perverse joy in leading around in circles.

Had Szilard been content simply to remain a critic on the sidelines, he might have been ignored by Groves. But what Szilard described as his lifelong commitment "to save the world" inevitably brought the two men into conflict. Szilard provoked Groves first by organizing a petition of atomic scientists to oppose the atomic bombing of Japan, and then by an irregular intercession with James Byrnes, President Truman's secretary of state designate, in an equally futile effort to keep the bomb from being used as a diplomatic club against the Russians after the war. Using security as a pretext, Groves effectively derailed Szilard's hopes of becoming the spokesman for the atomic scientists as well as their postwar agent provocateur.

While Lanouette does not make the case explicitly, Groves may have been doing Szilard a favour. Szilard was always more an artist than an architect, and it is difficult to imagine what useful role he could have played once atomic energy became the subject of national legislation and the responsibility of a burgeoning bureaucracy. Frustrated, at least temporarily, in his mission to save

the world, he instead turned his efforts to understanding it better, taking up the study of biology, both experimental (molecular and microbiology) and theoretical (problems of ageing and memory). His contributions in these fields were never as great as those in physics, although once again he played the part of facilitator by encouraging young biologists and attempting to find scientific answers to social questions.

But in 1962, unable to stay away indefinitely from saving the world, he created the Council for a Liveable World, the first political action committee for arms control, which played a decisive role in the election that year of Senator George McGovern and which remains an active lobby today. Reverting to type, Szilard angered the Kennedy administration when he insisted on conducting his own personal diplomacy with the Soviet leader Nikita Khrushchev. Similarly, he alienated many even among his supporters in the council by his decision to flee the United States for Switzerland at the height of the Cuban missile crisis — he was the first refugee of the Third World War, he grandly if mistakenly declared. Back home, finding that his petitions now went unsigned and his urgent entreaties were ignored, he turned, wisely, to a new form of expression: fiction. But, although no longer delivered in broadsides, his message was essentially unchanged from his crusading days in the Manhattan Project.

Szilard's lifelong mission to save the world ended short of attainment. After a near-miraculous recovery from bladder cancer, he died in his sleep of a heart attack at the age of 66 in 1964. One of the ironies of the man is that he will probably be best remembered by the public neither as a physicist nor biologist but as an inadvertent futurologist. His blunder during the missile crisis notwithstanding, Szilard's was a remarkably prescient vision — predicting not only the start of the Soviet-American nuclear arms race almost a year before the first atomic bomb was tested, but also, more than a generation ago, the end of that race in an exhausted stalemate by the late 1980s.

While protesting throughout his life that he "would rather have roots than wings", Szilard's own choice as well as his destiny cast him as an intellectual gadfly. Lanouette's exhaustively researched and artfully written account of one of the most underrated figures of the atomic age establishes Szilard as both a curmudgeon and a posthumously honoured prophet. □

Gregg Herken is in the Department of Space History, National Air and Space Museum, Smithsonian Institution, Washington DC 20560, USA.

Chemical conventions

Joseph Rotblat

Banning Chemical Weapons: By Hugh D. Crone. Cambridge University Press: 1993. Pp. 122. £27.95, \$39.95 (hbk); £8.95, \$12.95 (pbk).

AN event of great historical importance took place on 13–15 January of this year: at a ceremony in Paris, 130 nations signed a treaty banning the development, production and use of chemical weapons.

There have been many international arms-control treaties in the past few decades, but each had some inherent weakness. The 1968 Nuclear Non-Proliferation Treaty, generally thought to have been highly successful (despite the recent defection by North Korea), actually legitimizes the possession of nuclear weapons, even though by only five privileged states. The 1972 Biological Weapons Convention is all-embracing, but has no proper verification mechanism to ensure compliance. By contrast, the 1993 Chemical Weapons Convention requires all signatory states to destroy their existing stocks of chemical weapons and prohibits them from developing, producing, acquiring or stockpiling such weapons in any circumstances. It also contains a comprehensive verification regime to ensure compliance by routine monitoring and 'challenge inspection' (under which a signatory state can be required to demonstrate its adherence to the treaty) and a procedure for action in cases of non-compliance.

It is astonishing that such comprehensive prohibition was achieved for the most difficult of the weapons of mass destruction. Chemical weapons are the easiest, cheapest and quickest to manufacture from materials widely available and constantly used in daily life. Several recent events may account for the readiness to reach final agreement. One is the employment of chemical agents in the Iran–Iraq war, and the general feeling of revulsion about their use against the Kurds by President Saddam Hussein. Another is the end of the Cold War and the unprecedented unanimity of the United Nations Security Council in peacekeeping efforts.

But there is still a long way to go from the signing of the treaty to its implementation, and there are many obstacles to overcome. The problems facing the Organization for the Prohibition of Chemical Weapons, set up to oversee the operation, include the destruction of the chemical stockpiles, which will take many years and be very expensive (more

than the cost of production) and potentially dangerous to the environment; and the unease of the chemical industry about intrusive monitoring and commercial secrecy. For these and other reasons, there will still be many groups seeking a better understanding of the issues surrounding chemical weapons.

Their quest will be met to a large degree by Hugh Crone's *Banning Chemical Weapons*. A senior research scientist in Australia, with many years of experience of the subject, the author has managed to pack an amazing amount of information into a slim volume and present it in a very readable style: concise yet thorough, factual but also entertaining, with frequent flashes of dry humour.

In these days of rapid change, authors of books on topical subjects are bound to find themselves overtaken by events, but this book manages to be remarkably up to date (it was nevertheless written before the treaty was signed). The accounts of the ways chemical agents may harm the human body, the offensive capabilities of modern chemical warfare, and the defensive measures that can be taken against them, fulfil Crone's aim to inform the reader about the relevant facts. But I find it difficult to accept his assertion that the material is entirely scientific and technical (the book includes a discourse on the origin of aggression and an essay on arms control), or his contention that the role of the scientist ends with the presentation of facts, leaving the rest to diplomats and lawyers. This sort of argument is usually voiced by those who dislike disarmament measures. Such problems, they say, are of a purely political nature and not therefore a matter for scientists to discuss; but when on the political forum, the same people use the incompleteness of technological knowledge as an excuse for not reaching final agreement.

In reality there is no clear-cut division between the different aspects. There are clearly problems with the chemical weapons treaty (for example in the definition of a chemical agent and in the use of tear gases and herbicides) that require the continuous interaction of scientists, technologists, industrialists, lawyers and politicians. The best illustration of where these various groups come together is



Evolution of the gas mask — the first respirator (top left) issued to British troops in France in April 1915 after the German chlorine attack at Ypres. By late 1916, the protection had progressed to the "small box" respirator (bottom right). This series of pictures appears in *Porton Down* by G. B. Carter, which traces the 75 years of chemical and biological research at this UK defence establishment. HMSO, £9.95 (pbk).

the Pugwash Conferences, at which chemical warfare has been discussed since 1959. The topic has been pursued, often with the Stockholm International Peace Research Institute, in many symposia and workshops. It is generally acknowledged that such informal discussions contributed greatly to the success of the official negotiations on the Chemical Weapons Convention at the Geneva Conference on Disarmament.

An important aspect of the success of the treaty is that it can serve as a stimulus for an analogous treaty on nuclear weapons. An argument usually trotted out against the elimination of nuclear weapons is that they cannot be disinvited: even if all nuclear arsenals were destroyed through an international treaty, this would not stop a future aggressor from producing a new arsenal. The chemical weapons treaty knocked this argument on the head. As Crone points out, a nuclear treaty would be difficult to violate because nuclear fuels are well characterized, have few places of origin and require special handling, in sharp contrast to toxic chemicals which are of many different types and can be produced in many plants and handled simply. So the fact that a comprehensive treaty that bans chemical weapons has been achieved should encourage, and serve as a model for, a similar agreement on nuclear weapons. □

Joseph Rotblat, president of the Pugwash Conferences, is at 8 Asmara Road, London NW2 3ST, UK.

Science and trans-science

Alvin M. Weinberg

The Rise and Fall of Nuclearism: Fear and Faith as Determinants of the Arms Race. By Sheldon Ungar. *Penn State Press*: 1992. Pp. 214. \$32.50 (hbk), \$14.95 (pbk).

Containing the Atom: Nuclear Regulation in a Changing Environment, 1963–1971. By J. Samuel Walker. *University of California Press*: 1992. Pp. 533. \$50, £34.50.

ALTHOUGH both of these books are 'historical' accounts of nuclear development in the United States, they differ both in the segments of the development they cover and, more fundamentally, in their approach to, even their conception of, 'history'. Sheldon Ungar is a sociologist. He covers the nuclear arms race not so much as history but as a phenomenon to be interpreted in an arcane sociological idiom. J. Samuel Walker is the official historian of the US Nuclear Regulatory Commission (NRC). This second volume of the official history of the Nuclear Regulatory Commission* covers, in a strict, thoroughly documented fashion, the tortured events that led to the public's disaffection with nuclear energy and to the attempts by the Atomic Energy Commission to quell that disaffection.

Ungar's book poses serious difficulties for the reader, such as me, who is unfamiliar with the methods, and perhaps the pretensions, of modern sociology. He defines 'nuclearism', following Robert Lifton, as "the faith in awe-inspiring power" (of the atom) as a source of salvation — for example, the United States was able to guarantee the freedom of Western Europe by threatening a nuclear response to Soviet aggression.

Nuclear fear

But according to Ungar, this "numinous" aspect of the bomb was always accompanied by "nuclear fear": the bomb was both omnipotent and dread-inspiring. The various episodes in the nuclear arms race — missile gaps, arms limitation treaties, Atoms-for-Peace, the Cuban missile crisis, even the eventual ending of the Cold War — resulted, Ungar believes, from oscillations in the relative strengths of the polity's appreciation of one or the other aspect of the bomb. When confidence in the bomb's omnipotence dominated, US foreign policy, as in Vietnam, was aggressive. But when confronted with the reality of the bomb, as in the Cuban missile crisis, nuclear dread prevailed. The United

States, as well as the Soviet Union, was cautious, and the crisis was resolved.

Ultimately, according to Ungar, 'nuclearism' failed: nuclear dread made the bomb "unusable". Here Ungar quotes McGeorge Bundy, who, in his book *Danger and Survival: Choices About the Bomb in the First Fifty Years* (Random House, 1985), points out the "enormous gulf between what political leaders really think about nuclear weapons and what is assumed in calculations of relative advantage in simulated strategic warfare. . . . In the real world of real political leaders . . . even one hydrogen bomb on one city would be recognized . . . as a catastrophic blunder." Thus, as Tom Schelling first pointed out, and as Ungar implies, the fall of nuclearism suggests that the world may be creating a "tradition of non-use"; perhaps this tradition of non-use, if sufficiently strong, will forever prevent a nuclear holocaust.

Readers who wish to know just what happened during the heyday of nuclearism will find Bundy's *Danger and Survival* a far more lucid and better documented work than Ungar's book. For those who understand and are sympathetic to sociological interpretation of history, Ungar's book provides provocative, though unprovable, insights.

Containing the Atom is a marvellously informative account of how the US Atomic Energy Commission coped with its responsibility to regulate nuclear energy during Glenn Seaborg's chairmanship. Nuclear energy for electrical power production was just coming of age; but public apprehension over the safety of nuclear energy was also beginning to emerge at this time.

The commission was mandated by the 1954 Atomic Energy Act to produce nuclear weapons, to promote the private use of atomic energy and to ensure the safety of commercial nuclear power. To carry out the last mandate, the commission established a small regulatory staff. Even at the time, many observers realized that the commission's regulatory responsibility could be in conflict with its other two promotional responsibilities. Walker's tale of the often excruciating dilemma the commission faced in discharging its conflicting responsibilities reads like a novel — with heroes, villains, suspense and human frailties.

The difficulties first became apparent during the original fallout scare of the 1950s. At the heart of the issue was the biological effect of low levels of radiation: did exposures of the order of background radiation cause harm, as Linus Pauling claimed, or were these levels innocuous, as the commission in-

sisted? To a remarkable degree, the question of the effect of low levels of radiation was at the bottom of most of the controversies that embroiled the commission during the 1960s: waste disposal, reactor safety and, of course, radiation standards. Even in a reactor accident, the huge number of casualties predicted in the worst conceivable accident results from assuming a no-threshold linear hypothesis. But the question is trans-scientific: at levels around background, science cannot say what is safe, what is not.

Acceptable standards

How can the regulators set standards of acceptable exposure in such cases? Here I believe that the entire nuclear regulatory apparatus fell into a trap. Standards for the public were set as 1/10 (later 1/30) of the 0.05 sieverts per year established for radiation workers. This standard came from the experience of early X-ray workers and really had little bearing on chronic effects of radiation. Far better would have been to set radiation standards as a small fraction of the unavoidable background level (0.001 sieverts per year). Whatever harm, if any, is caused by exposure at this regulatory standard, it would be small compared with what everyone already accepts. The commission did eventually lower the permissible exposure to the public from a nuclear power plant to some 5×10^{-5} sieverts per year — but only after an excruciating, even humiliating, confrontation with congressman Chet Holifield, chairman of the Joint Committee on Atomic Energy.

How could the commissioners, honourable and able people, discharge these conflicting responsibilities in a way that would satisfy both promoters and detractors of nuclear energy? Obviously they could not satisfy everyone, perhaps even themselves. This was poignantly illustrated by the commission's suppression, in 1965, of the famous revision to report WASH-740. In this report, Brookhaven scientists had estimated that the worst possible reactor accident could contaminate as much as 100,000 square miles of land and cause damage of \$17 billion. Although commissioner John Palfrey urged publication of these findings, the rest of the commission objected — basically because publication would erode public support for nuclear power. In the end, Palfrey acceded to the majority's view.

In retrospect, the commission's decision was probably wrong. But the pressures on the commission by the Joint Committee and the nuclear industry, as well as their own nobly inspired desire to see some good, not only evil, from nuclear fission, almost like a Greek tragedy led them to their decision. And

*For a review of the first volume, see *Nature* **318**, 607 (1985).

so it went for each of the issues that the commission had to confront: mining, emergency core-cooling, waste disposal, thermal pollution, environmental impact — in every case the commission tried to balance its desire not to impede nuclear power with its responsibility to assure public health and safety. Although it is easy in retrospect to criticize the commission for being overly permissive in its regulation of nuclear energy, the results as seen in 1993 are proof that something was being done right. Despite the accident at Three-Mile Island, no member of the US public has been harmed by nuclear power.

That an official history of a US government agency pulls no punches, but gives the facts just as they were, is remarkable. Walker and the Nuclear Regulatory Commission are to be congratulated for providing us with so balanced and accurate an account of how nuclear regulatory policy was established in the United States. □

Alvin M. Weinberg, former director of Oak Ridge National Laboratory, is in the Medical Sciences Division, Oak Ridge Institute for Science and Education, PO Box 117, Oak Ridge, Tennessee 37831-0117, USA.

Ancient aqueducture

M. J. T. Lewis

Roman Aqueducts and Water Supply. By A. Trevor Hodge. Duckworth: 1992. Pp. 504. £55, \$78.

ROMAN aqueducts are a never-failing source of interest to archaeologists, engineers and laypeople alike. So it is astonishing that no general study has hitherto been published in English or (of anything like this quality) in any other language. Hodge triumphantly combines deep scholarship and lucid interpretation, while guiding nontechnical readers considerately and painlessly through the inevitable quagmire of hydraulic principles.

He follows the water over its whole course from intake to domestic plumbing, and finally sees it down the drain. En route he discusses the Romans' predecessors, wells and cisterns, surveying, irrigation, mills and baths, and urban distribution. The book is thorough, utterly practical and highly readable — even at times funny — and it will be a long time before it is superseded.

Our knowledge about the subject is limited to two principal literature sources. Frontinus's justly famous book on the aqueducts of Rome is informative about administration but thin on technology, and does not necessarily apply to other cities. Vitruvius, although trying to describe the practicalities, does not always succeed, to the point where one wonders if he knew what he was talking about. The great bulk of the evidence, then, is archaeological. The most obvious components — the arcades striding across the Campagna, the mighty bridges of Segovia and Tarragona or the Pont du Gard — are the least typical, for the aqueducts mostly ran underground in cut-and-cover trenches or in tunnels. With all their accessories such as dams, inverted siphons, settling tanks and water towers, aqueducts provide plenty of scope for debate about how they worked.

So, inevitably, not all specialists will agree with Hodge's conclusions. He emphasizes the risk, for instance, of mains pressure blowing the plug of domestic taps out of their hous-

ings, without considering the attempts to hold them down; and many would question his remarks that most Roman water-mills were supplied by aqueduct, or that in northern Europe the horizontal water-wheel predominated.

His subject is so large that some prime examples have found no place: Dorchester's open-channel aqueduct; Mérida's fine system of sewers; Frontinus's information on the maintenance staff at Rome. Hodge shirks the fascinating issue of Lincoln, where the aqueduct's source lay 30 metres below the water tower. And Norman Smith's recent paper on the Pont du Gard, although in the bibliography, is not discussed. Smith suggests that the topmost tier of small arches, a unique feature, was a last-minute addition to retrieve a disastrous error in levelling. With these arches, the aqueduct just worked, falling only 6 metres in 25 km between Gard and Nîmes, the minimum gradient being an extraordinary 7 cm per km; without the arches, the aqueduct would not flow. The idea is plausible, for Roman engineers, like any others, were not infallible.

Such omissions as these, however, are minimal when set against the whole. Many a hoary myth is laid to rest, most notably that the Romans avoided pressure systems because of the weakness of their pipes. On the contrary, Hodge clearly shows that when crossing a valley up to about 50 metres deep (as at the Pont du Gard), they used a bridge; if deeper, they used an inverted siphon. To cite the extreme instances, at Beaunant near Lyon the head was 123 m and the pressure 1,207 kPa (175 lb in⁻²), whereas on the Hellenistic siphon at Pergamon the figures were about 190 m and 1,825 kPa (265 lb in⁻²). So much for fear of pressure: these siphons were not only built with lead pipes, but worked.

With such achievements, few would deny that Roman aqueducts were a magnificent feat of engineering and organization. This, and the benefits they brought, Hodge fully acknowledges. Neither is he starry-eyed about them, for he sees them in context. Often they were status symbols, crippling to the council's budget, built to keep up with the municipal Joneses, supplying vast quantities of water to cities whose drains could not cope. Most urban dwellings, moreover, were not on the mains for reasons of cost or lack of head, and continued to rely on wells and rainwater cisterns just as they always had. Which goes to show that in no age or society does engineering, however prestigious and expensive, necessarily impart universal benefit. □

M. J. T. Lewis is in the Department of Adult Education, University of Hull, 49 Salmon Grove, Hull HU6 7SZ, UK.

Bernard Cox/Architectural Association

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Geomythology

L. Sprague de Camp

Arktos: The Polar Myth in Science, Symbolism and Nazi Survival. By Joscelyn Godwin. *Thames and Hudson/Phanes*: 1993. Pp. 260. £10.95, \$16.95 (pbk).

JONATHAN Swift, the misanthropic Anglo-Irish cleric (1667–1745), once spun a little verse that runs:

So geographers in Afric maps,
With savage pictures fill the gaps;
And o'er uninhabitable downs
Place elephants in place of towns.

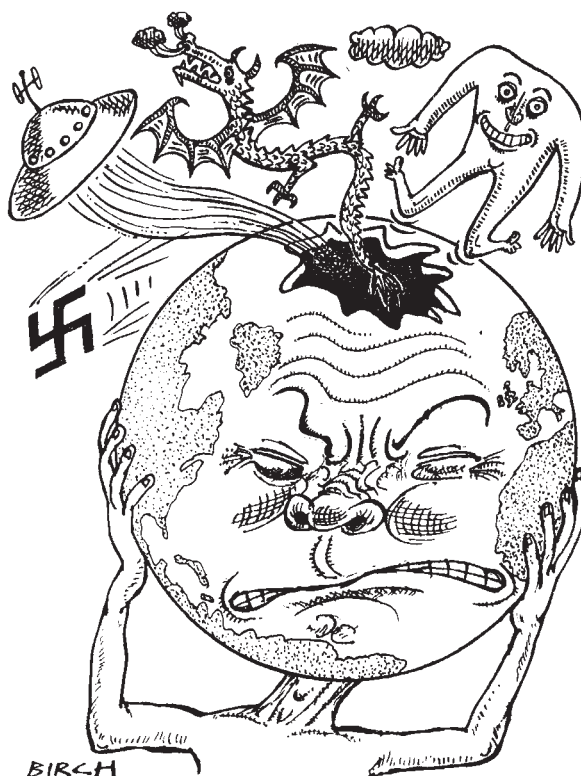
Back before the entire Earth had been, if not explored on foot by civilized men, at least flown over and mapped, geographers put things stranger than elephants beyond the boundaries of the known: one-eyed Arimaspians, one-legged Sciapodes, headless men with faces in their chests and fabulous wildlife. These animals were usually composites of known creatures. They included the unicorn of India, the gryphon of Scythia, and dragons everywhere. In 1952, my lamented friend Willy Ley and I attempted a popular survey of leading geomyths in our *Lands Beyond*, which Joscelyn Godwin cites as a source.

Godwin also surveys geomyths, focusing on those associated with the Arctic. He begins with the myth of the Golden Age, before men discovered civilization and fell into its evil ways. The three 'Abrahamic religions' (Judaism, Christianity and Islam) have their own version in the Garden of Eden.

The big, fat Russian adventurer who became the uncrowned Queen of Occultists, the fascinating Helena Petrovna Blavatsky (1831–91), founder of theosophy, made up her own Golden Age, located in the North Pole. According to her and her disciples, man evolves through seven Root Races, each with seven subraces. We are the fifth Root Race; the sixth will soon appear.

The first Root Race was a kind of astral jellyfish, living on the Imperishable Sacred Land centred on the North Pole. The animals are unknown as fossils for the obvious reason that jellyfish seldom leave fossils. After the Imperishable Sacred Land disappeared (despite its imperishability), the second Root Race, a little more substantial, dwelt in Hyperborea, roughly Siberia plus Scandinavia and the Canadian Arctic. The

third Root Race comprised the apelike, hermaphroditic, egg-laying Lemurians, inhabiting the large southern continent of Lemuria. Some had four arms; some, eyes in the backs of their heads. Their discovery of sex precipitated their downfall. (Blavatsky took a dim view of sex, at least after she became too old to be interested in it herself.) And so on through the wholly human Atlanteans to us.



Hyperborea, like Atlantis, comes from ancient Greek geographical speculations. The Hyperboreans were supposed to dwell in the far north, perhaps on the north coast of Asia. Never having been there, Greeks imagined the Arctic as a fine, balmy place, where men lived a thousand years.

Godwin's pursuit of northern lore brings in a vast mass of nineteenth-century science, speculation and fakery. Examples are the rise of the Aryan myth, embraced in the twentieth century by the Nazis with genocidal results. It involved the dispute of the 1840s over whether man's original language had been Hebrew, as Biblical literalists assumed, or Sanskrit, as the Aryanists declared.

Another such fragment sucked into

the occult and pseudoscientific vortex was the fictional force 'vril', which appeared in the science-fiction novel *The Coming Race* (1871) by the prolific English writer Edward Bulwer-Lytton, who a few years before had been made first Baron Lytton. In this tale, Lord Lytton takes his hero into an underground cave world. There he finds a race of supermen using vril, which they mentally direct to blast rocks and monsters. Blavatsky and her theosophists adopted vril as an effective means of propulsion and defence for her prehistoric races, like Edgar Rice Burroughs' Eighth Barsoomian Ray. There was even a Vril Society, which included Karl Haushofer, the German founder of 'geopolitics'.

And so on and on, through the Thulean Society, a subgroup of which became the German Workers' National Socialist (Nazi) Party; the Chilean occultist Miguel Serrano, who hailed Hitler as the Tenth Avatar of Vishnu and claimed that Hitler had escaped the fall of Berlin in a German flying saucer; the Polaires or Polar Society, which enlisted the occult help of the ghost of Sir Arthur Conan Doyle; jewel-crusted Shambhala, either on the surface of Tibet or Mongolia or beneath it, and ruled by a King of the World; and even the notorious Shaver hoax perpetrated by the late Raymond Palmer in 1945–48, when he edited *Amazing Stories*. Palmer received a number of manuscripts from the semiliterate Richard S. Shaver, asserting as fact that the world's ills were caused by evil telepathic emanations from a race of subterranean fiends, the deros. Palmer asserted that he, too, took these assertions as true. Visitors to *Amazing's* office were amazed to find Palmer cowering behind his typewriter and listening for deros.

Godwin goes on to serious discussion of the form of the Earth, especially of those who thought it was hollow. Eminent among these was John Cleves Symmes, who said that the Earth was not only hollow, with holes at the poles, but also that it was merely the outermost of a whole series of hollow spheres, one inside the other.

One group held not only that the Earth was hollow, but also that we were inside it. This group was headed by Cyrus Teed, who ran his cult on the west coast of Florida from 1894 to his death in 1909. When he failed to resurrect in three days as promised, several followers killed themselves; but stubborn believers lingered on for years.

Having myself read through much of the Higher Nonsense in getting material for my own books, I must admire Godwin's stamina in wading through the vast quantities listed in his bibliography. His attitude towards the prophets of Hidden Wisdom and Secret Masters is admirably fair and nonjudgemental, albeit when considered *en masse* these colourful and riotously contradictory doctrines merely lend proof to Bertrand Russell's statement that most people, if you shout "Two and two make five!" at them long and loudly enough, will come to believe that two and two *do*, after all, make five.

As to whether the leaders of the countless occult and magical cults believe their own claims, who knows? We do not yet have a mind-reading machine, and Godwin does not address the topic. Cyrus Teed probably believed his own line. Contrariwise, Blavatsky seems to have been altogether cynical. Finding that she could make people believe that two and two make five, she exploited this gift to the hilt, as the most effective way open to her to make a living.

In one of her impulsively candid moods, Blavatsky told the Unitarian preacher Conway: "It's all glamour. People think they see what they don't see; that's all there is to it." The Russian journalist Solovyoff (or Solovëv) quoted her thus: "What is one to do when, in order to rule men, you must deceive them, when, in order to catch them and make them pursue whatever it may be, it is necessary to promise and show them toys?" Some purveyors of the Higher Nonsense may have begun their leadership from pure self-interest, but as time passed they came to believe their own pitch. (This may have been the case with the late L. Ron Hubbard of Scientology fame.)

Although Godwin seems not to be taken in by the prophets of Hidden Wisdom and Secret Masters, his attitude towards "materialistic science" is hardly more favourable. He accuses scientists of "brainwashing" the young with their doctrines of boundless "progress". This is confusing the scientific beliefs to which accumulating evidence pushes scientists with their beliefs on nonscientific (social, moral or cultural) matters. There is no hard-and-fast connection between the two sets of convictions. Thus the great ethologist Konrad Lorenz revealed Nazi tendencies. Godwin wants to hang on to the "realities and truths that stand inviolate", such as those attributed to Christianity.

When much younger, I rejoiced in freedom from religious dogma, thinking: wouldn't it be fine if everyone adopted my own hard-nosed, sceptical, materialistic, nonsupernatural, scientific view? Now, having seen the results of trying to run an empire (the Soviet Union) on the

basis of a purely secular, nontheistic philosophy, I am not sure that masses of men can manage their affairs without supernatural beliefs about, for example, rewards and punishments after death, or promotion and downgrading in one's next incarnation. So let Godwin hang on to his Christianity — if he can — even though I might not accept his arguments myself.

The work is a remarkably comprehensive coverage of the pseudoscientific field in recent centuries, well written and knowledgeable. If the author wanders off the boundaries of polar geomorphology, that is inevitable in trying to make sense of such hazy, amorphous concepts. It is like trying to measure a cloud with a surveyor's tape. Nor is the subject trivial. Sixty years ago, many well-informed people dismissed the Aryan-race cult as crackpot nonsense, and we know that use was made of the myth. □

L. Sprague de Camp is at 3453 Hearst Castle Way, Plano, Texas 75025-3605, USA.

Out of the bell jar

Hugh Freeman

Touched With Fire: Manic-Depressive Illness and the Artistic Temperament. By Kay Redfield Jamison. *Free Press (Macmillan Inc., New York): 1993. Pp. 370. \$24.95.*

"ONE of the oldest and most persistent of cultural notions", says Kay Jamison, is that of a possible link between madness and genius, even though many of the outstandingly creative have always appeared to be healthy and well balanced. In her view, confusion about the meaning of 'madness', together with ignorance particularly of the nature of manic-depressive (or affective) illness, has been responsible for most of the controversy around this theme. It has also tended to fall between two academic stools — psychiatric research, which is mainly biological, having little interest in the experience of writers and artists, and scholars of creativity fruitlessly pursuing a connection with schizophrenia. Jamison, though, was not only the co-author of one of the standard works on affective disorder, but has personally investigated a large British sample of artistically creative people. With these unique qualifications, what has she made of the 'mad genius' notion?

Her declared purpose is "to make a literary, biographical and scientific argument for a compelling association between two temperaments — the artistic and the manic-depressive — and their

relationship to the rhythms of the natural world". At the same time, she warns that labelling anyone who is unusually creative or eccentric with this diagnosis "both diminishes the notion of individuality within the arts and trivialises a very serious illness". Nor are most people who have affective disorder unusually creative. For a start, though, there have to be clear distinctions between affective psychosis (at the extreme end of the spectrum), episodes of major depression (or, more rarely, of mania) and fluctuations of mood. If these disorders are confused, as they often are, it becomes impossible to compare any one set of data about them with another.

Should anyone be disinclined to take manic-depressive illness seriously, Jamison points out that, if untreated, its mortality — now mainly through suicide — is higher than that for many types of heart disease; out of a long list of poets, a staggering 18 per cent had taken their own lives. There may also be disguised cases: Byron's involvement in the Greek War of Independence may in fact have been a way of ensuring his own death. The purpose of knowing this, however, is not just as an exploration of psychopathology, but to understand the man and his work more deeply. Tabulating what is known about 37 British and Irish poets born between 1705 and 1805, Jamison records that no fewer than 30 of them had evidence of significant affective disorder.

While all this information is impressive, it involves what epidemiologists call the 'floating numerator' — a problem which Jamison does not deny. Not only is the denominator of 'poets' very difficult to define, but literary scholarship has resulted in our knowing much more about the famous than about the unknown, which could bias the evidence. An alternative approach, though, comes from affective disorder being "the most genetic of the major psychiatric illnesses", providing what Jamison describes as "the constitutional core of a determining temperament". Although the genetic mechanisms are still largely unknown, studies of separated twins provide strong support for the illness being familial, with many unaffected relatives also showing cyclical swings of mood. What is particularly relevant here is that unusual creativity and other abilities may run prominently through the same families.

So the disorder is "paradoxically advantageous and yet destructive". Its most characteristic feature is a cycle of fluctuations in mood, energy, sleep — and creativity in those who have the gift for it. Quite recently, 'seasonal affective disorder', with depression related to the darkness of winter, has been found to be relatively common, and the incidence of suicide peaks in spring and autumn.

Jamison suggests that manic-depressive illness is in fact the expression of "mal-functions in a master biological clock". Her studies show pronounced seasonal patterns in productivity among writers and artists, confirmed by the biographies of many outstanding ones (Schumann composed over 130 songs in one year). It is quite likely that these creative periods coincide with swings of mood — particularly upwards — and that these in turn may coincide with seasonal changes in natural light.

If Jamison's general thesis is true, serious philosophical and ethical issues are involved, particularly now that better treatments are available for affective disorder. Most sufferers seek treatment

today, even if it may dampen down their creativity, but those in the past, of course, had no such choice open to them. Jamison deals with the complex scientific issues in a vivid and readily intelligible way, only marred by occasional descent into *Time-Life* language and some diagnostic imperialism in assuming that the US system is the only one worth mentioning. These are minor blemishes, though, in an important work that should provoke some serious rethinking in several corners of academe. □

Hugh Freeman is editor of the British Journal of Psychiatry, 17 Belgrave Square, London SW1X 8PG, UK.

Our world in collision

Reinold Schröder

Universal Ice: Science and Ideology in the Nazi State. By Robert Bowen. *Belhaven: 1993. Pp. 189. £39.50. (US dist., St Martin's Press.)*

NOWADAYS the name of Hörbiger evokes only the celebrated Viennese theatrical dynasty, especially its founders, the brothers Paul and Attila. It is less likely to bring to mind their father, Hanns, and his "World Ice Doctrine". Not so in Germany 50 years ago.

Robert Bowen, a geologist who concluded his academic training in Münster in 1991, has sought to rescue Hörbiger's cosmogony from oblivion, and in particular to discover how it was that this abstruse farrago of pseudoscience found so many adherents in the Third Reich, reaching indeed to the uppermost pinnacles of power. For they included Adolf Hitler and above all Heinrich Himmler.

Hanns Hörbiger (1860–1931) was a successful Austrian engineer, to whom the World Ice Doctrine came, late in the nineteenth century, "in bitter anguish of the spirit, as a vision"; its "cosmic, fathomless depths" threw his body into "feverish, shuddering spasms". Out of this experience Hörbiger drew his theory, which linked terrestrial and cosmic processes in a unifying drama of nature. In 1913, working with the amateur astronomer Philipp Fauth, he set down his cogitations in a book entitled *Hörbiger's Glacial Cosmogony*. Bowen's heroic endeavour has been to assimilate this barely readable work, comprising close on 800 pages of impenetrable gibberish, and boil down its message to a few dozen comprehensible pages. He has avoided unfair selectivity without sacrificing the perspective of a trained scientist.

According to Hörbiger then, the Mil-

ky Way is made up of fragments of ice that lie outside the gravitational field of the Sun. When this veil of ice passes into the domain of the planetary system, it plunges into the Sun, melts and evaporates. Immense exploding funnels — the sunspots — eject mighty fountains of steam — the solar prominences — which instantly freeze to ice powder; this showers down on the inner planets, which thus acquire a thick glacial crust. The Moon itself has a coat of ice several kilometres thick. In the Earth's atmosphere the ice dust melts into rain, while larger lumps (meteors) burst and strike the Earth as hailstorms.

But it is not only ice that descends on the Sun and planets. The all-encompassing ether slows the movement of all the heavenly bodies until they are trapped by the gravitational pull of the nearest inner planet and become moons, later to be absorbed by their planet. The planets are gradually approaching the Sun in spiral trajectories and will finally be drawn in and consumed. Our own planet too once possessed other moons, which crashed into the Earth, thus provoking natural catastrophes such as the immersion of Atlantis and Noah's Flood.

It is not surprising that a theory that scorns so many established facts and rests confessedly on nothing more than intuition and myth was rejected, or worse ignored, by the world of science. Hörbiger saw himself indeed in the line of Julius Robert Mayer, discoverer of conservation of energy, another unrecognized genius.

After this promising start, Bowen's book is a sad disappointment, for it tells us nothing new about the part played by the World Ice Doctrine in the history of the Third Reich. Readers have first to fight their way through a rather disjointed potpourri of German intellectual

history in the 1920s and '30s. From Nietzsche on Wagner, the Mormons and the "Nazi philosopher" Heidegger, the text moves on to the campaign by the "Aryan physicists" (Lenard and Stark) against "Jewish physics" (quantum theory and relativity). The relation of all this to the World Ice Doctrine is left obscure; errors of fact accumulate (Fritz Strassmann, who worked with Otto Hahn on the discovery of nuclear fission, for instance, is given a posthumous Nobel prize), and standard works on the history of science in Germany at that time have not been cited, or, it seems, even consulted.

Nor do we learn how Hörbiger's theory influenced the Nazis. There are only a few pages in which it is related in detail that in 1937 a certain von Elmayer-Vestenbrugg, writing in a volume of the *Kampfschriften der Obersten SA-Führung* (Books of Struggle of the SA High Command) broke a lance for the World Ice Doctrine, commending this Nordic concept of the origins of the world and of natural catastrophes, so shamefully disregarded by the academics, as an appropriate Nazi ideol-



Hörbiger and detail from his *Glacial-Kosmogonie*

ogy. Here Bowen unjustly accuses Hörbiger of embracing the ideology of the Nordic master race, for which there is no evidence in his writings. Neither does he reveal whether the SA author's article met with a favourable reception, or which members of the Nazi leadership declared for Hörbiger in consequence.

There is contradictory evidence in the literature as to whether Hitler was a follower of Hörbiger. It is certain only that he did not exert himself to promote the dissemination of Hörbiger's ideas. For Reichsführer SS Himmler, it was quite another matter: the studies of Brigitte Nagel (*Die Welteislehre: Ihre Geschichte und ihre Rolle im "Dritten Reich"*, GNT, 1991) have exposed him as the highest-ranking supporter of the

World Ice Doctrine. He began a search for conclusive evidence in the SS "Research Institute" ("Ahnenerbe", or "Ancestral Heritage"), with a view to elevating the doctrine to official Nazi ideology.

It is true that Himmler was not deflected from his faith in the World Ice Doctrine by the vehement opposition of the academic establishment, the "journeymen of science", for whom he harboured a deep loathing. This is certainly to be inferred from his harsh treatment of opponents of the Hörbiger ideology in his own ambit. But he was compelled to give the impression of distancing himself from the World Ice Doctrine, and even the journals under his control were deterred from running advertisements for new books about the theory. The research groups in the Ancestral Heritage were renamed and were able to pursue their researches only covertly.

This fiasco resulted primarily from the activities of the physicist and Nobel laureate — and ardent admirer of Hitler — Philipp Lenard, who vigorously fought the World Ice Doctrine as "Volksverdummung" — dulling of the people's mind. One may then hazard the interpretation that Himmler sought to protect Werner Heisenberg — who had been libellously dubbed "the Ossietzky of physics" by Johannes Stark, Lenard's spiritual patron, in the *Schwarze Korps*, the journal of the SS — as a means of striking at his arch-enemy in the propagation of the World Ice Doctrine, Lenard, whom he did not dare to challenge in matters of science.

Bowen is seemingly unacquainted with the state of knowledge in this area, for he finally goes so far as to say that "Hörbiger might well have become a scientific overlord in Nazi Germany probably more important than was Trofim D. Lysenko in Stalin's Soviet Union". In fact, even if Hörbiger had been alive, the World Ice Doctrine would never have achieved the standing of an official Nazi ideology. Such an outcome was not even attained by Lenard's "Aryan physics". In the course of the war the pragmatic professionals gained the upper hand over the pure Nazi ideologues in all important matters relating to military technology, and thus furthered the implementation of Hitler's destructive mania. And Hitler's paranoid hatred of the Jews did not stem from the World Ice Doctrine, but rather from the racial theories thrown up by social Darwinism at the turn of the century. □

Reinold Schröder is at Vaihinger Strasse 37, 7000 Stuttgart 80, Germany.

■ Nature thanks Walter Gratzner for translating this review from the German.

Invisible Johnny

Ray Monk

John von Neumann. By Norman Macrae. *Pantheon: 1992. Pp. 405. \$25.*

DURING his all-too-brief life (he was 53 when he died), John von Neumann was regarded by many, including both prominent politicians and Nobel-prizewinning scientists, as the most intelligent person in the world. That he is not better known among the general public is prob-

in connection with the 'von Neumann architecture' that lies at the heart of all modern computers.

Like that of those other great mathematical logicians, Kurt Gödel and Alan Turing, von Neumann's role in shaping the development of the twentieth century was, until recently, known to only a few. Public awareness of the other two, however, has been massively increased by the publication of two widely read books: Douglas Hofstadter's astonishing bestseller *Gödel, Escher, Bach* (Basic Books/Penguin, 1980) and Andrew Hodges's superb biography *Alan Turing:*

The Enigma of Intelligence (Simon and Schuster/Hutchinson, 1984). The success of these two books, both of them rich in scholarship and exciting to read, should have paved the way for a study of von Neumann that was lively enough to excite, and well-informed enough to satisfy, the curiosity of a wide audience about the nature and the importance of his achievements. Sadly, this is not that study.

Norman Macrae is a former editor of *The Economist*. He writes in a style that, while accessible, is perhaps too breezy for his subject matter. Von Neumann is called "Johnny" throughout the book, which is bad enough, but worse is Macrae's habit of putting an adjective in front of this familiarized form of his subject's name and referring to him as, for example, "cherubic Johnny", "inexperienced Johnny", "Computer Johnny", "logical Johnny" and, towards the end of von Neumann's life, "overweight

Johnny" (although, thank God, he spares us "cancerous Johnny"). Breezy Norm, one feels, does not read his work aloud. How else can one explain his letting go unrevised a phrase like: "the appearance of Johnnies is the cheapest way of increasing man's material prosperity very fast"?

Even worse than these stylistic infelicities, however, is the fact that Macrae does not seem very interested in mathematical logic. His work is therefore uncertain and unreliable at precisely the point when it should be laying an intelligible and enlightening foundation for what is to follow. For, as with Alan Turing, von Neumann's later multifarious activities were all in some way grounded in his early work in logic.

Indeed, one feels throughout the early sections of the book that Macrae is just dying to get on to what he is really interested in: von Neumann's work on economics, computers and nuclear deterrence during and immediately after

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

John von Neumann with his wife, Klara, and pet dog Inverse (1954).

ably due to the fact that, unlike say Einstein or Freud, his name is not associated with a single revolutionary breakthrough in any one discipline. His fame is, rather, diffused throughout the astonishingly varied fields of his interests. Mathematical logicians know him as the man who, in the 1920s, provided one of the first systems of axioms for set theory; in theoretical physics he is known as the mathematician who gave precise form, in a well-known book of 1932, to the so-called 'Copenhagen interpretation' of quantum mechanics; social scientists will have come across his name as one of the joint authors, together with Oskar Morgenstern, of the classic work of game theory; historians of the Cold War will know him as the technical expert responsible for advising first Truman and then Eisenhower on the strategy of 'maximum retaliation'; while computer scientists — and, increasingly, the general public — will recognize his name

the Second World War. His chapter on the Budapest in which von Neumann grew up, for example — potentially a fascinating and lively study of a pre-First World War culture as intriguing as its Habsburg counterpart, Vienna — leans far too heavily on statistics (“By 1910 Jews made up around 60% of Budapest’s doctors and lawyers”) and makes little attempt to breathe life into these facts or to use them to conjure up in the reader’s imagination a world that, though now largely forgotten, produced, in addition to von Neumann, Theodore von Kármán, Michael Polanyi, Eugene Wigner, Edward Teller and Leo Szilard.

In connection with this, one should also point out that the general standard of scholarship throughout the book is not very high. In places, Macrae makes heavy use of work that has already been published, and although in his notes he makes clear which particular books he is dependent on, he very rarely gives the exact source of his quotations or of his information. This is a pity, because where I am in a position to check his information, the number of mistakes he makes does little to inspire confidence in the rest. He has three or four pages on Bertrand Russell, for example, in which almost every paragraph contains an egregious error. Russell’s grandfather, Lord John Russell, for instance — a radical Foxite Whig for whom anti-Toryism was practically a religion — is described by Macrae as “a right wing former British prime minister” and a “promising young conservative politician at the time of Waterloo”.

There must, then, be doubts about how far we can trust Macrae for an assessment of von Neumann’s work, especially in the field of politics. His prejudice in favour of a hawkish, right-wing opposition to communism is plain for all to see, as is his willingness to distort the views of his opponents. Nevertheless, in part because of Macrae’s personal commitment, the sections of the book dealing with von Neumann’s work on the ‘Fat Man’ bomb that was dropped on Nagasaki and his role in first developing the hydrogen bomb and then advising Eisenhower’s government on the deployment of thermonuclear weapons are unfailingly fascinating. In dealing with the politics and personalities of committees, Macrae shows the confidence and the sure touch that was missing in his earlier accounts of mathematical logic.

What made Hodges’s book on Turing such a stunning success was the author’s ability to move around between three apparently unrelated subjects: the persecution of homosexuals in British society, the cracking of the ‘Enigma’ code and the development of computer models of the mind. A successful biog-

raphy of von Neumann would need to show a similar, perhaps even greater, range of interests. In this deeply flawed work, Macrae has done little more than hint at what such a book would be like. □

Ray Monk is in the Department of Philosophy, University of Southampton, Highfield, Southampton SO9 5NH, UK.

■ Of related interest is *Prisoner’s Dilemma: John von Neumann, Game Theory and the Puzzle of the Bomb* by William Poundstone, now published in paperback by Oxford University Press (£7.99) and Anchor/Doubleday (\$12.95). For a review by Martin Shubik, see *Nature* **356**, 637 (1992).

Have book, will travel

Karl Sigmund

Reality Rules: Picturing the World in Mathematics. Volume 1: The Fundamentals. Volume 2: The Frontier. By John L. Casti. Wiley: 1992. Pp. 388/424. \$80, £42.50 (two-volume set).

STARTING with the Necker-cube-like ambiguity of its title and going right up to a discussion of alternative interpretations of quantum reality in the final chapter, *Reality Rules* invites one continuously to shift one’s position. This plea for mobility is characteristic of John Casti’s book. It is essentially a sightseer’s guide to the fairyland of mathematical models, and aims at enticing one to book for a tour.

Casti Tours offers the most spectacular vistas of modern applied mathematics — including (of course) chaos theory, fractals, computational complexity, NP-completeness, artificial life and cellular automata, as well as such topics as catastrophe theory and game theory that are no longer the current rage but still provide reliable, well-tested thrills. Even the more down-to-earth topics on the agenda, such as linear systems and control theory, are replete with ‘souped-up’ applications ranging from Searle’s ‘Chinese room’ test for artificial intelligence to a party-goer’s optimal drinking policy. With the exception of a chapter on connective structures, which seems somewhat at odds with the other chapters, the common thread running through the book is dynamical systems in one guise or another.

The book grew out of *Alternative Realities*, a widely acclaimed work published in 1989 that was awarded a best book prize by the American Association of Publishers. In the meantime the book

has grown by 300 pages and split into two handsome volumes. The main changes include a substantial new chapter on computation and complexity, a wealth of exercises, problems and discussion questions, a solutions manual available on request and a wholesale updating of the text.

The book’s style can best be seen by considering the author’s antecedents. Casti started out as a systems scientist at the Rand Corporation, and wrote several mathematical textbooks before launching into science writing for the general public with his well-received books *Paradigms Lost* (Morrow, 1989) and *Searching for Certainty* (Morrow, 1990) — reviewed in *Nature* **385**, 293–4; 1992). *Reality Rules* occupies a niche somewhere between a textbook and a trade book. It is intended for readers with some training in undergraduate mathematics and a curiosity about the latest ‘bestseller themes’ in the scientific arena. Not incidentally, Casti nowadays works at the Santa Fe Institute, a hotbed of avant garde activities in dynamic systems modelling.

The two chapters bracketing the book are general discussions about the aims and limitations of mathematical modelling. Each of the eight chapters in between outlines enough material for at least a one-term graduate course. These surveys are to a large extent independent of one another (so that the same topics are sometimes approached from several perspectives), but they are not meant to be self-contained. On the contrary: in a sense, the climax of each chapter is reached with the six to ten superb pages of notes and references, which cover the relevant literature in an informal, agreeably chatty and extremely user-friendly way. I found them most helpful, both in fields I am familiar with and in those areas where I am a stranger. They offer reliable, thoroughly up-to-date advice for mathematical globe-trotting.

The text is designed to whet one’s wanderlust. It is mostly a blend of simple yet instructive examples and heroic theorems. The general impact of these theorems is then analysed in a conversational style that leads up to a collection of stimulating discussion questions. Casti’s formula is thus to invite the reader to take some easy steps first and then offer a breath-taking view of the heights. The notion of an algorithm, for instance, is explained by means of a detailed recipe for Caesar’s salad, but just a few pages later one is up to one’s ears in uncountable numbers. In much the same vein, the exercises are a mixed bag — some straightforward classroom drill problems together with many thoroughly hard nuts that on occasion summarize entire research articles, sending one scurrying to the solutions manual. These

abrupt shifts in gear can be upsetting, especially as they are hidden by Casti's easy style, which can cause one to exceed the speed limit. The book deserves to be taken at leisure. Breezing too rapidly through *Reality Rules*, one might get the impression that it really reels.

Like most travel guides, the book does not recoil from hype; but it is used in a relaxed way, secure in the confidence that if readers don't buy one cruise, they'll book another. This nonchalance reflects the current trend away from theories toward models. A theory can have supporters and opponents, each committed to the hilt; but a model cannot cause fanaticism — you're free to take it or leave it in the tolerant eclecticism of the Roman Pantheon. Casti is no button-holing zealot out to convince. In fact, his lines are studded (artfully, I suspect) with casual asides meant to provoke even the most docile of readers

into occasional mutterings, containing such off-handed remarks as that Western science arose because the 'competition' was unable to come up with a satisfactory explanation of the Black Death, and claiming that some innocuous incidence matrix yields 'insight' into a Shakespearean sonnet. But these are minor eddies in a broad stream; the tale runs smoothly on, and carries such a wealth of information that one doesn't feel like quibbling over matters of opinion with such an urbane cicerone. As a racy synopsis covering a huge area, *Reality Rules* will best be enjoyed by scientists who have retained not just a working knowledge of college mathematics, but also a youthful taste for the excitement of heady intellectual adventure. □

Karl Sigmund is in the Institute for Mathematics, University of Vienna, Strudlhofgasse 4, A-1090 Vienna, Austria.

Symmetry for the armchair

Simon L. Altmann

Symmetry in Chaos: A Search for Pattern in Mathematics, Art and Nature. By Michael Field and Martin Golubitsky. Oxford University Press: 1992. Pp. 218. £19.95, \$35.

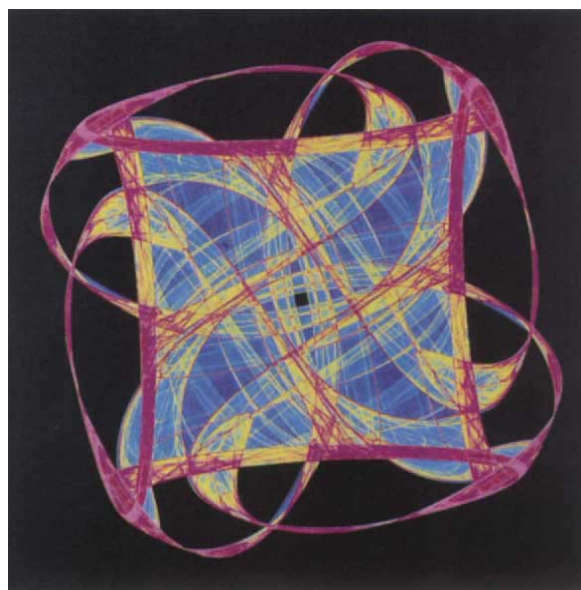
Reflections on Symmetry in Chemistry... and Elsewhere. By Edgar Heilbronner and Jack D. Dunitz. VCH: 1993. Pp. 154. DM58, £22.

The Dynamics of Ambiguity. By Giuseppe Caglioti. Springer: 1992. Pp. 170. DM98, £40.

EVER since Hermann Weyl published his famous *Symmetry* in 1952 many have been called to write on the subject; few, however, are those who can move, like Weyl, with equal authority from Hans Castorp to the problem of inversion in Raphael's cartoons. In *Symmetry in Chaos*, Michael Field and Martin Golubitsky, quite wisely, have given themselves a different remit and have succeeded in producing a book that is in itself a work of art: not a minor achievement for a book on mathematics. They have in fact invented a wonderful scheme to produce two-dimensional coloured patterns by computer.

If you allow your cursor to perform a random walk, pixel by pixel, on a computer screen and if you have a colour scale in which the pixel colour is determined by the number of visits that it has received, you will of course get, after sufficient iteration, a uniformly coloured rectangle. Field and Golubitsky, instead, superimpose on this process an attractor, in the sense of chaos theory. This is an iteration scheme that will guide the ac-

tive point with greater probability towards some particular areas of the screen and that is cleverly designed to lead to one of the two-dimensional symmetry groups (C_4 in the illustration). The authors provide sufficient explana-



Swirling Streamers taken from *Symmetry in Chaos*.

tion of the mathematics to make their method understandable, as well as the ideas behind it, and full programs in BASIC are given in the appendices. As the authors themselves say, the question whether these pictures, and the underlying mathematics, will be of lasting scientific interest "is one that has yet to be answered". These two mathematicians, on the other hand, have now

joined the ranks of the great inventors of two-dimensional patterns, from the craftsmen of the Alhambra to the modern ceramic, textile and wallpaper designers.

A general book on symmetry that does not quote or reproduce M. C. Escher is a rare one these days and Field and Golubitsky's is no exception to the rule. But in *Reflections on Symmetry in Chemistry... and Elsewhere*, Edgar Heilbronner and Jack D. Dunitz go further, because they are the first people I know of to point out that patterns in which the background is a translated mirror image of the foreground were in fact invented by Kolo Moser decades before Escher. Anyone fortunate enough to gain access to the recently refurbished (but alas still erratically open) Museum für Angewandte Kunst in Vienna can see a cabinet by Moser with a marquetry frieze where the positive and negative layers are identical after reflection (in practice, a 180-degree rotation about an axis on the layer) and translation. It is possible that the challenge to achieve this in marquetry drove Moser's invention, but he also used the idea in textile design.

This detail in Heilbronner and Dunitz's book is typical of the meticulous care with which it has been written. Without any mathematics, the more important ideas arising from group theory are given and beautifully illustrated. To achieve such a degree of informed simplicity is indeed most difficult and the authors even succeed in explaining sufficiently the concept of irreducible representation so as to be able to discuss Woodward and Hoffmann's rules. But there is a great deal more in this excellent work, from very illuminating symmetry games to historical discussions on the use of symmetry principles. This is a book that one can warmly recommend to intelligent sixth-formers and undergraduates alike.

Such an audience, however, cannot easily be imagined for *The Dynamics of Ambiguity* by Giuseppe Caglioti, now translated from the Italian by A. O. Bucci. But none of its readers will fail to be struck by the width of the author's interest in broken symmetries, from quantum mechanics, music and art analysis à la Arnheim, to Virgil's poetry. □

Simon L. Altmann is at Brasenose College, Oxford OX1 4AJ, UK.

Science and psychosis

John C. Marshall

The Cognitive Neuropsychology of Schizophrenia. By Christopher D. Frith. Lawrence Erlbaum: 1993. Pp. 169. £14.95, \$28.50.

"I FEEL a piercing stare from behind. I feel people want to harm me but I will not fight and my enemy will be disarmed." The writer is Vaslav Nijinsky, *Le Dieu de la Danse*, retreating into madness as the greater insanity of the 'Great' War lurched to an exhausted halt. In the epilogue to the diary that he wrote in St Moritz during 1918-19, Nijinsky noted that "the doctors do not understand my illness", a reflection that prompts the question whether, 75 years later, our present understanding of schizophrenia is materially better. Christopher Frith's monograph is a superb trampoline on which to bounce this question, for it provides both a succinct summary of what is known (or conjectured) and a highly original approach to the notion of explanation in psychiatric disorders.

As Frith points out, "schizophrenia is a surprisingly common illness with a life-time risk of approximately 1 in 100 people". The 'cost' of schizophrenia in terms of suffering is incalculable, but because the only values our society now accepts are based on market economics, Frith also reminds us that the cost "in terms of treatment, care, and skills lost to the community, is at least 2% of the gross national product". Yet despite the obvious magnitude of the problem, medical science cannot yet diagnose schizophrenia.

Lest this remark sound like mere wilful obfuscation, I should elaborate the grounds on which it is based. "Diagnosis [I quote the *Oxford Companion to Medicine*] is the process of identifying the disease or other circumstances responsible for the patient's complaints." No such disease has (yet) been identified as 'schizophrenia'. Rather, there are various classification schemes, based on check-lists of signs and symptoms that allow the physician to label patients as 'schizophrenic'. But the fact that psychiatrists can employ such check-lists reliably and consistently is a far cry from discovering the underlying disorder (or disorders) responsible for the symptomatology.

These check-lists are furthermore grossly heterogeneous. The most commonly employed scheme (DSM-III-R) includes the characterization that schizophrenic psychosis must include two of the following symptoms: delusions, prominent hallucinations, incoherence, catatonic behaviour, flat or grossly in-

appropriate affect. Whether these symptoms form a theoretically coherent group (or even a set of family resemblances) remains in doubt. The classification also contains the exclusionary criterion that there should be "no evidence of organic factors" involved. As Frith wryly notes, a truly bizarre paradox results: it is "widely believed that there is an organic basis to schizophrenia", but "a patient

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Nijinsky as Petrushka.

can only be diagnosed schizophrenic as long as the organic cause of the illness is unknown!"

In the first part of his monograph, Frith attempts to cut the Gordian knot by seeking explanations not of schizophrenia but rather of particular schizophrenic symptoms. Consider the main symptom of vocal hallucinations (hearing voices). Frith provides convincing experimental evidence against the notion that patients misperceive (and hence misinterpret) external auditory stimulation, and the beginnings of evidence for the notion that patients fail to recognize their own subvocal speech as their own. He then attempts to show how such failures of "self-monitoring" could in turn lead to a belief in (for example) "thought insertion" (where patients "experience thoughts coming into their mind from an outside source") or "delu-

sions of control" (where "patients experience their actions as being controlled by an outside force").

In the concluding chapter, Frith then looks at both a *tour de force* and a volte-face in which schizophrenic disorders of willed action, of self-monitoring and of monitoring the intentions of others are all brought together as a unitary impairment of the ability to form "metarepresentations". In autism, the child "does not try to infer the mental states of others"; in adult schizophrenia, the patients misrepresent their own and other people's propositional attitudes to the content of their thought. This section fits poorly with the rest of the book and betrays evidence of "over-inclusive" thinking, as Frith himself slyly lets slip. The notion is so striking, however, that one hopes that Frith will develop the idea with the same close attention to fact and logic demonstrated in the other chapters.

Throughout those chapters, Frith links particular schizophrenic signs and symptoms with malfunction of particular brain regions that are known (from studies of other pathologies) to be responsible for specific cognitive processes. These brain systems constitute a corticostriatal loop whose functioning is modulated by the neurotransmitter dopamine. Drugs that block dopamine receptors remain the most effective treatment for the florid positive symptoms of acute schizophrenia. But, as Frith notes, their exact mode of operation remains controversial and the linking hypotheses that mediate between biology and behaviour need to be spelled out in considerably more detail than is currently available. "Over-sensitivity to dopamine" does not in itself constitute an explanation of vocal hallucinations and paranoid delusions.

There is also, sadly, evidence that some of the cognitive impairments seen in chronic 'burnt-out' schizophrenia are "a necessary consequence of the mechanism by which dopamine blockade reduces positive symptoms". These side-effects include reduction of spontaneous activity and severe difficulties with concentration, although a similar deterioration was often seen in the days before the introduction of antipsychotic drugs. The ubiquitous Harry, Count Kessler, was shocked to encounter Nijinsky in the foyer of the Paris Opéra in December 1928: "His face, so often radiant as a god's, for thousands an unforgettable experience, was now gray, flabby, empty, only fleetingly lit by a blank smile, a brief gleam as of a flickering flame. Not a word crossed his lips." □

John C. Marshall is in the Neuropsychology Unit, University Department of Clinical Neurology, Radcliffe Infirmary, Oxford OX2 6HE, UK.

How to skin an elephant

John A. J. Gowlett

Making Silent Stones Speak: Human Evolution and the Dawn of Technology.

By Kathy D. Schick and Nicholas Toth. Simon and Schuster: 1993. Pp. 351. \$25.

LOUIS Leakey, famed for his discoveries at Olduvai Gorge, was also one of the first to bring the stone age to life with dramatic experimental studies. Two American researchers who have greatly developed this tradition are Kathy Schick and Nick Toth. Experimenting and writing two academic generations on from Leakey, they pay generous tribute to their own inspiring teachers at Berkeley in a lively book that aims to evaluate the role of technology in human evolution. This is a timely emphasis: as archaeology strove to give itself a social face in the 1970s and '80s, technology was often overlooked, largely because earlier studies had been excessively typological; but now there is a broader reappraisal.

There is nothing dull in the approach taken by Schick and Toth, who hardly hesitate even to skin an elephant with stone flakes. They do admit that "the sight of a twelve-thousand-pound animal carcass the size of a Winnebago can be quite intimidating", but in a bold experiment small lava flakes proved up to the task. Such work illustrates one of the niches that may have been open to tool-using hominids, who were probably able to cut into carcasses whose tough hides held other carnivores at bay. Their studies have led the authors to a respectful interpretation of the early hominids, whose activities show considerable foresight — "not a simple story of the most expedient technology".

Much of the book is a straightforward and up-to-date account of human evolution, but the authors have found several ways of enlivening it. They use italicized chapter introductions to give an 'early-hominids-eye' view of the world, pithy quotations to reinforce their points, and numerous action photographs of their own experiments. Each author makes a substantial personal contribution in terms of experimental archaeology. Schick is mainly concerned with how early sites were formed, including the element of human behaviour that can be isolated behind many natural factors. Toth is the replicator of stone artefacts, and thus a discoverer of the routines that govern their manufacture. These complementary approaches give a refreshingly balanced view of human evolution,

with special insights where the authors' own work is concerned.

One such area is early hominid use of the landscape. Schick has developed the late Glynn Isaac's food-sharing model into a more neutral 'favoured place' hypothesis. This involves hominids returning frequently to places that offer both good food supplies and suitable protection from predators. Another remarkably exciting line of enquiry concerns the abilities of tool-using chimpanzees. To operate their comparative framework, anthropologists have to evaluate any overlaps in ape and human behaviour. Since 1990 the authors have collaborated with Sue Savage Rumbaugh in exploring just how far a chimp can go. A certain Kanzi proves able to appreciate sharp cutting edges and to strike flakes from stones. Chimpanzees are the only primate species apart from humans that regularly make tools. Some recent work has suggested that their ability to do this is on a par with early hominid technology, but this research seems more likely to pinpoint the distinctions that exist. Kanzi, it turns out, is principally employed in communication studies, "so he could only engage in these experiments from time to time". He is evidently more in de-

mand than the poor old robust australopithecine of Toth's lament, consigned to the fringes of human evolution:

I got the Zinjanthropus blues
My name's been abused
They call me dumb, they call me thick
My tools don't even do the trick, oh no
I just can't take it no more
I may just do myself in
Down on the FLK floor

Palaeoanthropology is a controversial and often hotly contested field, but there is no edge to this book, other than that of its stone tools. Schick and Toth's pleasure in their work shines through, an enthusiasm that is infectious. They have great confidence that technology is an inescapable part of our heritage, and no scepticism about its role in the present and future. When those silent stones are made to speak, they seem to tell not just of the past but also of those who study them now — like the early hominids, Schick and Toth have taken up stone tools and taken on the world. In this fascinating account they show that stone tools do not just endure, but also have an enduring relevance. □

John A. J. Gowlett is in the Department of Archaeology, University of Liverpool, Liverpool L69 3BX, UK.

The tangled bank revisited

G. A. Dover

Genes in Ecology. Edited by R. J. Berry, T. J. Crawford and G. M. Hewitt. Blackwell Scientific: 1992. Pp. 534. £50, \$109.95 (hbk); £25, \$52.95 (pbk).

IN ending his most famous advocacy for evolution by natural selection on a lyrical but pertinent note of "... the tangled bank... clothed with varied plants... birds singing... insects flitting... worms crawling... dependent on each other in so complex a manner...", Darwin firmly tied evolution to its ecological roots. For the ensuing theoretical high priests of evolution spiralling off into other-worldliness, tangled maths replaced tangled banks. Happily, a reversal in attitude is unfolding, following hard on the heels of new molecular and ecological insights that are wresting the study of evolution from its 'Heath Robinson' theoretical contraptions and placing it back firmly in life, as we recognize it, on Earth.

Why did ecologists abandon evolution to the geneticists? Was it because of the perceived difference in timescale between ecological and evolutionary dynamics or because the near-theological battles waged by population geneticists over selection versus neutrality, and other

huffs and puffs, left them bemused and generally disenchanted? Ecologists view life as more than balanced polymorphisms, the myth of constant selective values, the selfish-genery of big *A* versus little *a*, just-so adaptationism and stable equilibria of large single-species populations harmoniously occupying ready-made niches. Ecologists were down on their knees in the tangled bank, rooting for cause-and-effect, dissecting the chaotic fluctuations of multi-species interactions, deconstructing the confounding duality of the organism-niche phenomenon, and decomposing the idealized population into its metapopulation substructure; whilst not forgetting for one moment that the individual organism, in all its prime and glory, was the central unit of evolution determining the subsequent fate of each of its thousands of fractious genes. If ever two disciplines were destined to pass each other like ships in the night, they were ecology and evolution, notwithstanding the joint kickstart they got from Darwin.

I exaggerate (of course) in order to polarize the extremes and to give credence to the thrust behind this unusual book and last year's meeting of opposites, in Norwich, England, from which it sprang. The editors rightly judged that

the time was ripe to knock ecological and genetical heads together for the grand reckoning, to force distinguished strangers to present and write joint papers, breaking down what J. Turner points out as 'cognitive particularism' in science: the inability on the part of one group to conceive that what the opposition is saying makes any sense whatsoever! By this device, no one individual could hide for long behind the comfort of esoteric jargon or enjoy the luxury of personally imploding on a microspecialization. Refreshingly, we are treated to honest dialogue, some of it still smouldering from fires not quite extinguished (J. Turner: "My own view is that there is no good evidence in favour of the neutral theory"), but all with the sense of urgency that the two disciplines need to merge, with more than a backward glance to the approaching charge of the new breed of molecular ecology imperialists (well represented here by M. Kreitman, P. Young and T. Burke).

A *Who's Who* of leading investigators spiritedly lay out their working philosophies with R. May, B. Clarke, B. Shorrocks, S. Stearns, P. Harvey, B. Levin, M. Begon, M. Hassell, J. Endler, R. Koehn and W. Watt and many other friends and foes taking as much interest in each other's perspectives as they can currently muster. One major gap in the listings is the absence of house-trained molecular biologists with first-hand experience of the fluidity of the genome and its contributions *per se* to standing levels of molecular diversity at the population level. Many molecular probes are not well-behaved mendelian elements obligingly evolving in a clock-like manner: they have a life of their own which we ignore at our peril.

W. Provine, A. Cain, S. Berry and A. Bradshaw provide the historical background, showing that blame for the schism that almost isolated the study of evolution from the college of scientific disciplines has to be laid at the door of some influential theoretical geneticists. Wrapped within the *a priori* belief in universal selective determinism they fought against neutrality and the molecular clock, constraints, trends and stasis, sympatric speciation, genetic revolutions, chaos, nonadaptive evolution, the stochasticity of metapopulations, probability refuges, hierarchies of selection, reaction norms, exaptations, suggestive directed mutagenesis, and (I would add) the ubiquitous mechanisms of genomic flux that underpin evolution by means of molecular drive. Wielding their equations and algorithms, they flay all such phenomena either to marginalize them or tame them within the high moral ground of panselectionist theory. "Ecology is what organisms do when they are doing nothing interesting" might sound

amusing from the mouth of E. B. Ford but is disastrous as a scientific attitude (realized or not) in those who assumed the mantle of Darwin.

There is a constructive place for theory in all sciences but if evolutionary theory is to be brought back to Earth, then a deep re-examination of its assumptions and parameters is in order; "How actually" must replace "how possibly" (as R. Sibly and J. Antonovics point out) no matter how tangled, complex, irrational and haphazard real organisms are in their real environments. M. Feder and W. Watt, in particular,

provide a comprehensive and persuasive pointer on how to reach this "Holy Grail".

Evolutionary theory needs to move on from naive caricature and application of darwinian selection in the same way that physics moved on from newtonian mechanics. By embracing the internal dynamics of the genome and the external dynamics of the ecology, there is hope for it yet. □

G. A. Dover is in the Department of Genetics, University of Leicester, Leicester LE1 7RH, UK.

Imperfect unities of knowledge

John Ziman

Common Sense, Science and Scepticism: A Historical Introduction to the Theory of Knowledge. By Alan Musgrave. Cambridge University Press: 1993. Pp. 310. £35, \$54.95 (hbk); £12.95, \$17.95 (pbk).

Reading the Book of Nature: An Introduction to the Philosophy of Science. By Peter Kosso. Cambridge University Press: 1992. Pp. 198. £22.95, \$34.95 (hbk); £7.95, \$10.95 (pbk).

The Disorder of Things: Metaphysical Foundations of the Disunity of Science. By John Dupré. Harvard University Press: 1993. Pp. 308. \$35, £27.95.

THE 'Two Cultures' thesis of the 1960s challenged the narrowness of science education. So science students were coaxed into courses on the history and philosophy of science as a natural bridge into the humanities and as a background to their own specialized studies. This didn't work.

One reason was that history and philosophy of science, as normally taught, turned out to be just as specialized and dry-as-dust as any scientific discipline. Another was that it all seemed very remote from what the students were learning to do in their science courses. Instead of illuminating scientific practice, philosophy merely made it seem vaguely questionable.

The first two of these books are typical products of this educational fiasco. Alan Musgrave does the conventional round of gurus — Aristotle, Descartes, Kant *et al.* — and their epistemological doctrines — scepticism, empiricism, rationalism and so on. For no obvious reason he then takes the popperian exit, which he labels "fallibilist realism". Surely, for a science student, that's like Omar Khayyam in his youth, going out "by that same door as I went in".

On the other wing, it's all 'common sense', with no damned nonsense about

rigour. That's obvious from Peter Kosso's title, which he keeps repeating as if he meant it. In the seventeenth century people genuinely believed in a benevolent, omniscient Author. Nowadays, we describe scientists as *writing* books about nature rather than *reading* a pre-ordained text. Tush!

With the philosophy of knowledge in such doldrums, monocultured scientists do better to concentrate on the philosophy of nature. Instead of asking questions such as 'how is it we know?' and 'how do we know it?', it is more fruitful to ask 'what is it we know?' and 'what do we know about it?'. John Dupré's book is original, lucid and confident, without being eccentric, polemical or arrogant. It deserves close attention.

To the average scientist, especially on the physical side, the unity of science seems indisputable. But it is clearly a metaphysical principle, beyond scientific verification. Dupré's thesis is both an assertion of how the world is and a denial that science constitutes, or could ever constitute, a single unified project.

His central target is obviously reductionism. Dupré takes this to mean the belief that the world can be classified hierarchically, and that the laws operating at higher levels can be derived for those at lower levels. If this were completely valid, then the world described or explored by science would be united by the laws of the lowest level. This is why some theoretical physicists look forward confidently to the discovery of a 'Theory of Everything'. But even if complete physicalism could never rule in practice, piecemeal reductionism is often presented as desirable, and feasible in principle, at every interface between levels — molecule to atom, cell to molecule, organism to cell and so on.

Dupré disputes this on deep-rooted theoretical grounds. He is not talking about the practical limitations of insufficient computer power, inadequate knowl-

edge of the relevant properties or our recent recognition of the characteristically chaotic phenomena that can arise even in deterministic systems. His argument is complex, with many subtleties to explore. But it seems to me that it is built around the following point, which he illustrates by an ecological example.

Suppose that we are looking for the scientific laws governing the respective abundances of hares and lynxes, and wish to 'reduce' these to the laws describing the behaviour of individual animals. We know that lynxes depend on hares for their food, and have certain capabilities for hunting, killing and eating them. We may also know, from very detailed observation, about the strategies used by hares to avoid capture, their breeding habits and so on. Nevertheless, no amount of this 'lower level' knowledge would be sufficient to determine the probability of capture in such encounters. This is a statistical property of the hare-lynx system, describing the outcomes of innumerable such encounters. Although it is functionally dependent on the behavioural characteristics of the species, it is defined only as the result of observations averaged over the populations in question. The reductionist project fails. The required property of individuals is not meaningful except in the context of scientific discourse at the population level. So it cannot be used to rewrite this discourse in the scientific terminology of individual behaviour. Ecology is not 'just' zoology on a larger scale.

Many important philosophical themes intersect in this focal area. Dupré first considers 'essentialism', as it arises typically in biological systematics. For example, are there certain essential features differentiating one species from another? What distinguishes hares from rabbits, which are also food for lynxes? Do the features used by taxonomists have any relevance to their capacity for escaping capture? Ecologically, should they be lumped together as a single population or treated separately?

It is often thought that sex is an essential property of an individual organism: should the males and females of these various species be considered as distinct natural kinds when it comes to hunting or being hunted? Dupré argues persuasively that sexual and gender differences do not support ethnographic, ecological or ethological generalizations, in humans or other animals. This is a socio-political insight that he rightly underlines in his final chapter on science and values.

Physicists and chemists are often taught, most unwisely, that biological taxonomy is akin to stamp collecting. It is a foundation law of physics that all electrons are identical; in chemistry, all



Prickly problem — "No doubt their spines are soft at the time of their birth, or else the poor dam would have but a bad time of it in the critical moment of parturition", wrote Gilbert White of the hedgehog in his *Natural History of Selborne*. This picture by Thomas Pennant (1726–98) appears in a new edition of this classic book. Thames and Hudson, £12.95.

nuclei of a given isotope are strictly interchangeable. The forthcoming 'Grand Unified Theory' will claim to describe 'everything' in terms of absolutely essential properties such as mass, charge and quark colour. Dupré might note that the operating context of the 'lowest level' has been expensively and lovingly contrived to evince precisely these properties. There is no empirical support for the belief that this downward analysis can be reversed to synthesize upwards through biology, psychology and human history.

That would require an account of all the causes that generate phenomena. Dupré shows that causality for the events at a particular level cannot be attributed exclusively to entities and events at the next lower level. Even the notion of 'levels' here is suspect. Suppose we are trying to prove that the behaviour of an organism is 'caused' by what goes on among its cells. How does this apply to a molecular pheromone causing sexual behaviour in a social situation? Ambiguities of categorization reduce the reductionist account to a shambles of improvisation.

The notion of 'probabilistic causality' is no help. In fact, as Dupré remarks "our intuitions about causality . . . are derived wholly from the macrolevel" — and yet when we want to give a complete causal story we are usually forced to retreat to the microlevel, where quantum uncertainty vitiates those intuitions. Speaking of causality, reflect on the thought that "what is problematic about humans is not that they are exceptions to an otherwise seamless web of causal connection, but that they are extraordinarily dense concentrations of causal capacity in a world in which such order is in short supply". Such insights are potent weapons.

Note that this is not mystical obscurantism. There is no appeal to supernatural, immaterial influences. It is not a debunking, relativizing attack on the philosophical ethos of the scientific enterprise. Apart from some justified scorn at the unrealized pretensions of certain disciplines, such as population genetics and macro-economics, Dupré accepts, indeed celebrates, many of the achievements of science. He approves of unifying developments that link previously disconnected parts of science or extend the range of theories or techniques.

All he asks for is a clear realization that reductive analysis is essential for an understanding of how a kind of phenomenon is possible, but tells us little about the actual course of events. For that reason "reductionism is a local condition of scientific research, not an irresistible tide sweeping the whole of science into an increasingly orderly pattern".

Dupré insists that there is no general scientific method, process or attitude. Science should be thought of as "a loose and heterogeneous collection of more or less successful investigative practices . . . varying from the highly plausible to the quite incredible, and from the extremely valuable to the entirely pernicious". He pins down the notion of the unity of science as a form of scientism appropriate only to a Utopia or to totalitarianism. He notes that "paradoxically, with the disunity of science comes a kind of unity of knowledge". That is why, to my mind, this is just the kind of philosophical teaching that is needed to close the gap between the two cultures. Here, surely, is the door through which we should, when young, have come out. □

John Ziman is at 27 Little London Green, Oakley, Aylesbury, Buckinghamshire HP18 9QL, UK.

The doctor's dilemma

David Weatherall

The Cambridge World History of Human Disease. Edited by Kenneth F. Kiple. Cambridge University Press: 1993. Pp. 1,176. £75, \$150.

THE *Concise Oxford Dictionary* defines a handbook as "a short manual or guide book". Usage has certainly changed over the years. Between 1883 and 1886 the New Sydenham Society published Charles Crichton's translation of August Hirsch's monumental two-volume *Handbuch der Historisch-geographischen Pathologie*. This remarkable work, which first appeared between 1860 and 1864, attempted to catalogue everything that was then known about the historical, pathological, social and geographical aspects of disease. Not surprisingly, no-one tried to repeat this *tour de force*, that is until about eight years ago when Cambridge University Press and a large international editorial board decided to attempt this daunting task. The product of the long gestation is this extremely large and handsome offspring of Hirsch.

As a co-editor of a textbook of medicine, which, though it will soon reappear in three volumes, covers only a small fraction of the field encompassed by the *Cambridge World History*, I was particularly interested to see how such a vast work would be organized. And how would the editors avoid those embarrassing letters reminding them of their sins of omission, in our case ranging from such dreadful gaffes as hiccuph and priapism to musicogenic epilepsy, a condition which, though it does not pose a considerable public health problem, means a lot to those who suffer an unpleasant spasm each time the national anthem is played? All of us have our pet disease and, not unreasonably, do not expect it to be neglected in a *Who's Who* of pathology.

The *Cambridge World History*, like the English hymnal, wanders freely between the ancient and modern. After brief essays on the history of Western, Chinese and Islamic medicine, there are accounts of changing concepts of disease, medical specialities, preventative medicine and ways of measuring health. The story is rounded off by an extensive description of the history of disease in different parts of the world and the present-day geography of disease, and by 158 essays on individual diseases, both current concepts and historical perspectives. Surprisingly, and thanks to an ingenious two-part 75-page index, it is quite easy for readers to find their way round this vast maze of pathology.

Of course, the editors had one large problem that did not face Hirsch in the

nineteenth century: what to omit? In some key sections a few examples have had to suffice. To illustrate how concepts of disease have changed, heart disease, mental illness, cancer and sexual deviation are chosen. This works fairly well, but the range of subjects used to illustrate medical specialities and disease



Doctor in the house — this woodcut appears on the title page of *Pliny et al.'s Aureum opus* (Pavia, 1516) and is reproduced from *Medicine and the Five Senses*, a collection of essays edited by W. F. Byrum and R. Porter. CUP, £40, \$59.95.

prevention is more idiosyncratic. Genetic disease, paediatrics and tobacco-related disease, here renamed tobaccosis, are understandable, but given the rich range of possibilities under these headings it is not clear why chiropractic and addiction come so high on the list. On the other hand, the extensive accounts of the history of human disease in different parts of the world include some excellent essays and, like those on the present-day geography of disease, contain material that would be difficult to find anywhere else summarized in such a compact and scholarly way.

The selection of the 158 diseases that form the subjects of individual essays must have been a nightmare for the editors. In the end they settled for a preponderance of infectious diseases and several disorders that have disappeared with time, including chlorosis, that strange variety of anaemia so loved by medical historians. But why ainhum and not appendicitis; and catarrh, diarrhoea and dropsy, but not that endemic scourge of Westernization and memorial to the arctic plumbing arrangements of British public schools and ancient universities, constipation?

Was it all worth it? With some reservations I think it was. Because of the huge arena covered, this is not a reference book for historians or medical specialists in their narrow fields. They are bound to find errors and omissions. Having a particular interest in genetic and haematological diseases, I was not happy about the balance and content of either of these subjects. And when I looked up my personal pet disease, thalassaemia, the commonest inherited disorder in man, I was dismayed to read that "it is most frequent in areas where ancient Greek immigration was most intense", a rather dated treatment of a condition that is now known to have

arisen by different mutations all over the tropical world and which affects hundreds of thousands of children in India and South-East Asia.

But the first edition of a work of this scope must inevitably lack balance in parts and contain some inaccuracies. Viewed in a broader perspective it represents a considerable achievement that brings together in one volume many of the diverse roots of modern medical practice and hence makes this absorbing story available to a much wider audience than has hitherto been possible. At a time when both historians and practitioners of science and medicine are concentrating on small and highly specialized areas of their subjects, and when undergraduate and postgraduate curricula are increasingly overcrowded, general reviews of this kind are welcome. Provided that a few of the problems of fact and balance can be ironed out in future editions, this new work could become a worthy successor to Hirsch's remarkable effort. □

David Weatherall is in the Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK.

Past as prologue

Ryan J. Huxtable

The Development of American Pharmacology: John J. Abel and the Shaping of a Discipline. By John Parascandola. John Hopkins University Press: 1992. Pp. 224. \$32.50, £27.

THE first citation in the *Oxford English Dictionary* for the word 'pharmacology' is dated to 1721. It was more than 150 years, however, before anything approaching modern conceptions of the discipline arose out of the conjunction and transmutation of *materia medica*, physiology and chemistry. This chimaera of a book — part biography and part history of science — traces the path by which spores from the late Victorian intellectual ferment in Europe were blown to the New World to seed the vigorous growth in pharmacology that we have seen this century.

The path was a strange one. Throughout the eighteenth and nineteenth centuries, Europe was abubble with provocative new findings. These ranged from the work of Scheele on oxygen in Sweden; Goethe's ultimately discredited ideas on colour and Sertürmer's isolation of morphine in Germany; the multifaceted investigations of Pasteur in France of crystal forms and his refutation of abiogenesis; the toxicological experiments of Francesco Redi and Felice Fontana in Italy; and the self-dosing with nutmeg and belladonna of Purkinje in the part of the Austro-Hungarian Empire that later became Czechoslovakia. Yet pharmacology in the United States descends in a straight line from none of these countries, but from the remote and bleak littoral of the eastern Baltic. Here, in the fens of Estonia, is the town of Tartu, formerly Dorpat. And there at the university in the 1850s, the professor of *materia medica* was Rudolph Buchheim. Through his textbooks and students, he can be claimed as the father of German and the great-grandfather of American pharmacology. He seems to have been the first to argue that pharmacology was a distinct discipline. One of his students was the Latvian, Oswald Schmiedeberg, co-founder in 1873 of the *Naunyn-Schmiedeberg's Archives of Pharmacology*, still in publication. Schmiedeberg left Dorpat to become professor of pharmacology at Strassburg, which was part of the German Reich from 1871 to 1918.

It was to Schmiedeberg's laboratory that the American, John Jacob Abel, came in 1886, beginning a pilgrimage that became a *sine qua non* for American pharmacologists until well into this century. For a considerable period, ev-

ery new chair of pharmacology in the United States was occupied by a student of Schmiedeberg. The influence he had in general is indicated by the fact that when he died in 1921, 40 chairs of pharmacology world-wide were occupied by his former students, a record that probably remains unbroken. Schmiedeberg, in continuation of the ideas of his mentor, distinguished *materia medica*, pharmacology and therapeutics, arguing that pharmacology provided the basis for the latter, but was a distinct discipline. His concept is enshrined in the name of the American Society for Pharmacology and Experimental Therapeutics, less formally known as ASPET.

Abel returned to the United States, first to the University of Michigan, then to take the first chair of pharmacology in the country, at Johns Hopkins University in 1893. He became the first president of ASPET and was the founder (in 1909) and first editor of the *Journal of Pharmacology and Experimental Therapeutics*. He also organized the American Society for Biological Chemistry and, with Christian Herter, helped to start the *Journal of Biological Chemistry*. At Johns Hopkins, Abel introduced laboratory work in pharmacology for medical students, one human demonstration being the effect of smoking on blood pressure. Such an approach was applauded in the Flexner report of 1910, which recommended that medical students should be exposed to laboratory work with animals in order to understand the experimental basis of therapeutics. Such exposure has now largely disappeared from the education of medical students, multiple-choice examinations having displaced laboratory work. Twenty years ago, students in my medical school had demonstrations on cats, dogs, frogs, guinea pigs, turtles and humans. Now only the latter species is used, in the emergency room on a Saturday night.

Another German tradition introduced by Abel has also disappeared. All his students were postdoctoral fellows. He opposed the awarding of doctorate degrees in pharmacology, believing that the ideal training for the aspiring pharmacologist was in chemistry and physics. So strong was the imprint of Abel on Johns Hopkins that it was not until 1969 that a PhD programme in pharmacology was established at the university. Now it is a rare pharmacology student who knows what happens when an acid is added to a base, or when sodium bicarbonate is heated. Students, however, have become adept at unwrapping molecular-biology kits.

Abel stressed the importance of daily get-togethers, where research problems could be discussed. Such gatherings were perhaps the most enjoyable and most

educational part of my own training as a chemist. Again, however, this is a tradition that has vanished from modern departments with their entrepreneurial laboratories.

Despite these changes, American pharmacology, to a considerable degree, still reflects the traditions established by Abel. His view of science, and its essentially holistic nature, is expressed by his remark to Flexner: "The investigator is a man whose inner life is free in the best sense of the word. In short, there should be in research work a cultural character, an artistic quality, elements that give to painting, music and poetry their high place in the life of man."

It is ironic that in his own research Abel is probably best remembered for his studies on adrenaline (epinephrine). In fact, he mangled the isolation of this unstable and readily oxidized substance. His procedures included treatment of tissue with trichloroacetic acid, and treatment of his extract with 25 per cent sulphuric acid or 15 per cent potassium hydroxide, procedures guaranteed to destroy substances more stable than a catechol. It is Jokichi Takamine who deserves most of the credit for establishing the structure of adrenaline.

It is important for pharmacologists to understand the roots of their discipline. Few do. But pharmacology, more than any other biomedical science, became what it is because of accidents of history. That is why the study of neurotransmitters is considered to be pharmacology rather than physiology, for example. If what's past is prologue, we must know the past in order to grasp the future. This is a period when pharmacology is under attack in the United States from parsimonious deans and curriculum committees that consider that the absence of the words 'molecular' or 'neuro' from the name of a discipline relegates it to the dusty museum of science, along with Felice Fontana's wax models and the jars of pickled specimens. By understanding why and how pharmacology arose, we are better able to defend the contributions that this flexible discipline still has to make to biomedical research and teaching. □

Ryan J. Huxtable is in the Department of Pharmacology, College of Medicine, University of Arizona, Tucson, Arizona 85724, USA.

New Journals

Nature's next review supplement is New Journals, which will appear in the issue of 7 October 1993. This annual survey will cover publications that first appeared during or after June 1991 and which issued at least four separate numbers by the end of April 1993. Further details will be given in next week's issue.

Fragrant aristocrats

Leslie Crombie

The Scent of Orchids: Olfactory and Chemical Investigations. By Roman Kaiser. Elsevier: 1993. Pp. 264. \$175, DFL280.

To most people who buy orchids in the florist's shop it will come as a surprise that a whole book has been written on the subject of these flowers' perfume. This is because we are accustomed to scentless or weakly scented commercial hybrids. Of 2,200 natural species examined, it is estimated that in fact 50 per cent are strongly or moderately scented, with only 15–20 per cent scentless to the human nose (and even these may not be scentless to an insect pollinator). The monocotyledonous orchid family is the most recent, yet the most extensive, class of flowering plant, with more than 25,000 species subdivided into 750 genera. To most of us, orchids are a class apart, aristocrats of the flower world. No other family excels these plants in range of colour, form and perfume. Roman Kaiser has brought together in this superb book his knowledge and enthusiasm for the botany and ecology of orchids, his skills as a photographer — most of the 170 orchid photographs are his — and his professional expertise as a chemist in the Swiss company Givaudan-Roure, where he has been investigating flower scents in relation to commercial perfumery for many years.

The book is divided into three parts, the first dealing with general information on plant scents, their role in pollination, the way in which orchid scents are collected and analysed without causing damage to the plant, and systems of verbal description of scents. There follows an interdisciplinary account of orchid species, their geography, and flower forms and scents, illustrated by excellent photographs. Finally, there are tables giving the analytical composition of orchid perfumes as determined by head-space analysis using capillary gas chromatography and mass spectrometry.

Scent from orchids is collected by enclosing the flower, passing over it a stream of air, and 'scrubbing' the air with a charcoal filter. According to the circumstances, anything from 1×10^{-6} to 3×10^{-4} g of scent can be obtained by elution from the column with solvent. Gas chromatography separates the components and about 1×10^{-9} g of each peak fraction must enter the mass spectrometer to get a suitable spectrum.

Mass spectrometry is the most sensitive instrumental detector, yet it is fascinating that the chemist with his nose at the end of a chromatographic column can detect (and sometimes identify) compounds that are in amounts too small to reveal themselves in the mass spectrum. The limits of detection of the human nose are about 10^{-12} g per litre for a strong odour, and some other animals have much more sensitive organs.

The ecology of orchid scent is also fascinating. In orchids that are moth-pollinated, scent is released only at night and complex symbiotic relationships between orchid and insect have been closely honed by parallel evolution. Aristocratic or not, some orchids are prepared to resort to downright fraud. *Draacula chestertonii* gives off a mushroom scent and the unusually large lip of the orchid imitates the cap of the

mushroom. This induces the mushroom fly to lay eggs while collecting or delivering pollen. When the eggs hatch, the larvae die of starvation through lack of mushroom nutrient. There is no symbiosis here.

The phrase 'coffee-table book' has a slightly derogatory ring and I hesitate to use it for this volume, but its many colour plates reach the best standards of such books. In addition, however, there is a lot of serious interdisciplinary text. So much care has gone into the book's production that it seems almost churlish to point to the occasional slip, as in the formula on page 127. *The Scent of Orchids* will stand as a landmark book on orchid perfume, while being a pleasure to read. □

Leslie Crombie is in the Department of Chemistry, University of Nottingham, Nottingham NG7 2RD, UK.

Correcting tunnel vision

Peter Harman

The Fontana History of Chemistry. By William H. Brock. Fontana: 1992. Pp. 744. £8.99 (pbk).

The Fontana History of the Environmental Sciences. By Peter J. Bowler. Fontana: 1992. Pp. 634. £8.99 (pbk).

THE publication of these accounts of the histories of chemistry and the environmental sciences launches Fontana's new history of science series*. On the evidence of these two volumes, the series will provide accessible yet authoritative surveys, from antiquity to the twentieth century, capped by valuable bibliographic essays that will be of great use to university teachers. The decision to take the story up to the present is especially to be welcomed, and will make the series relevant to students of the sciences; all too often, courses on the history of science are truncated around 1900 because of the paucity of suitable books.

In his preface to the two volumes, the series editor Roy Porter suggests that they should be seen as echoing current perspectives among historians of science, avoiding the anachronism of "tunnel history" (tracing narrowly defined topics through time). The decision to offer volumes devoted to the histories of different scientific fields, spanning more than 2,000 years of historical development, will however strike some historians of science as boldly toying with the pitfalls of anachronism. The problem is clearly understood and confronted by both au-

thors: in what sense do 'chemistry' and the 'environmental sciences' have a continuous history since antiquity?

In the case of chemistry, concerned with the properties and reactions of different kinds of matter, the problem is in conceiving the confluence of the various tributaries — metallurgy, pharmacy, dyestuffs, Greek philosophy, alchemy. The status of alchemy especially has long seemed troublesome, forming a decidedly controversial progenitor to the modern science of chemistry. In his recent book *Ideas in Chemistry* (which I reviewed in *Nature* 359, 785; 1992), David Knight argues that chemistry has had different and varied characteristics and suggests that its history should be investigated by a study of the ideas that have shaped its development and character. Although Knight's perspective on chemical history is programmatic and lacks the density of Brock's text, his point of view should be borne in mind, for Brock proceeds less circumspectly.

In the case of the environmental sciences, the problem, as Bowler notes, is that the history of the 'environmental sciences' may seem to be an "artificial category"; in other words, the definition and unity of the environmental sciences is "not created by the sciences themselves". Here there is the danger then that a very modern perspective — where a science such as ecology has been named only in the past century — seeks to group under a single umbrella modes of scientific inquiry that historically have no unity at all. The answer to the problem surely lies in understanding 'nature' in relationship to different cultural

*The series is to be published in the United States by Norton: *The Environmental Sciences and Chemistry* are issued next month, each at \$35 (hbk), \$15.95 (pbk).

frameworks, as Bowler himself suggests. But he does not carry through this approach systematically, so there is no clear thematic thread running through the volume. The bulk of the book is a fairly conventional (but informed and highly competent) survey of the history of geology, fossils and natural history, culminating in evolution and the darwinian revolution. The notion of the environment and its conceptualization in scientific studies lurks in the background, emerging only in the twentieth-century as the 'environmental sciences'.

The most striking feature of Brock's text is his focus on the period of modern chemistry following the 'chemical revolution' of Lavoisier (and others) in the late eighteenth century. Fully half of the book is given over to an excellent account of the principal developments of the nineteenth century: the chemical atomic theory, organic analysis, classification, structural organic chemistry, the periodic table, the origins of physical chemistry (in the chemistry of solutions), with glances at industrial chemistry, chemical education and publication. The distinctive feature of his approach is the very limited attention given to conceptual confusions over vitalism and organic models, in favour of an emphasis on the more technical questions of analysis, the concerns of the laboratory chemist. The account of twentieth-century developments is especially interesting, and though the treatment is selective — here we find discussions of Pauling, Robinson, Ingold and Woodward — it is presented in especially racy style. The concerns of Brock's text and his method of exposition (with necessary deployment of a limited chemical vocabulary) will make his book an excellent course book for history of chemistry courses taught in chemistry departments.

Bowler's text is much more suitable for use in general history of science courses. His synthesis of geology, fossils and evolutionary history is very well done; the breadth and wide chronological range of the book make it unique. The account of the most recent developments is particularly imaginative, and sparks off some critical engagements with controversial ethical questions surrounding the environment. Here the long historical perspective is especially valuable, pointing up the need to avoid simplistic constructions translated into political slogans and party programmes. Brock too concludes his book with a glance at the environmental implications of chemistry. And both authors are keen to use historical awareness to promote philosophical reflection. □

Peter Harman is in the Department of History, University of Lancaster, Lancaster LA1 4YG, UK.

Missing links

Alison Jolly

Visions of Caliban: On Chimpanzees and People. By Dale Peterson and Jane Goodall. Houghton Mifflin: 1993. Pp. 367. \$22.75.

VISIONS of Caliban is beautifully written, easily read and ethically challenging — it just might become primatology's *Silent Spring*. It confronts us with the plight of chimpanzees in Africa and the West. Although the stories are horrifying, they are told with relative restraint. Peterson



Condemned creature?

and Goodall (who write antiphonally) do not insist that all apes should be free — they even single out a few captive primate centres for praise. Rather, their central question is this: if biomedical research can benefit from the great apes precisely because they so closely resemble us, do we not have an obligation to treat apes according to their quasi-human emotional needs? Extremists who raid laboratories to lay hands on research animals can be fought by locks and guards and secrecy; but those who argue as cogently as Peterson and Goodall can be answered only by clear explanation of what laboratory apes can contribute to human welfare, whether this justifies our imprisoning them and — if it does — what conditions we owe them in captivity.

"History is marked by an amnesia about human origins and dependence. This great forgetting might be compared to an adult's powerful amnesia about infancy and early childhood", writes Peterson. As Gillian Beer has pointed out, the missing link can remain forever

denied, with its "freight of curiosity, reluctance, and disgust". This allows some of us the gusto of Alfred Russell Wallace when he shot a monkey in the Amazon: "The poor little animal was not quite dead, and its cries, its innocent-looking countenance and delicate little hands were quite child-like. Having often heard how good monkey was, I took it home, and had it cut up and fried for breakfast".

Peterson counters our amnesia by choosing Caliban as his missing link. This creature from *The Tempest* is part of Western mythology. We accept him as a half human, half beast, enslaved by the magician Prospero yet acknowledged as his kin. Similarly, we value chimpanzees because they are close to humans, yet treat them as animals to be used at our will.

Goodall and Peterson give us an up-to-date overview of research on the intelligence of chimpanzees. Chimpanzees, like Caliban, know "every fertile inch o'th'island": they have mental maps of the location of fruit trees and even know and use medicinal herbs. They hunt cooperatively for colobus monkeys and make tools that differ from one chimpanzee culture to another. Indeed, according to W. C. McGrew, some artefacts would be unattributable to humans or chimpanzees if they lost their museum labels. Chimpanzees have a sense of self, recognizing their own image in a mirror. They can learn the rudiments of language and understand novel sentences in spoken English if taught by a patient Prospero. Shakespeare was not far off when Caliban cries, "Thou taught'st me language, and my profit on't is, I know how to curse!" One of the common innovations of language-trained apes is the use of the sign for 'faeces' as an insult.

The heart of the book is the story of the 20 chimpanzees shipped by Franz Sitter from Sierra Leone in West Africa to Immuno Inc. in Austria. (Sierra Leone held roughly 2,000 wild chimps at the time, and perhaps ten apes die for each one that reaches a foreign country.) The authors reprint Shirley McGreal's exposé in the *Journal of Medical Primatology* in 1983. This letter caused Immuno to sue — each for \$4 million — McGreal, the journal's editor, publisher and distributor, and, incidentally, *New Scientist*, which had quoted criticisms of the company. Five years of litigation later, the Supreme Court of New York

threw out the case on the grounds that although McGreal's letter "might well have been damaging to Immuno, the facts she stated were 'evidently true'... [and] did not justify a libel suit". Meanwhile, all but one of the original defendants had settled out of court. It would make a great movie — but those cast as villains are still around.

There is more: the accounts, with names, of those who own chimpanzees as pets or use them as comic entertainers. All these people insist that they love the animals they keep. For some this is clearly true, for others, less so. But what emerges even more clearly is that no adult ape is a trustworthy pet — a chimpanzee is far stronger than a man, and has its own lusts and violent outbursts. One of the happiest captive chimpanzees may be Mr Jiggs, a female who works at birthday parties and bar mitzvahs with her jaws wired shut and an electric prod strapped to her back. The fate of a chimpanzee that is shut in a cage without social contact is far worse.

Goodall has sometimes been criticized for campaigning to help chimpanzees in captivity rather than addressing the destruction of wild populations; Peterson, however, has interviewed African hunters, traders and breeders and has located grisly trophies in the Brazzaville market. Of 25 countries that recently held wild chimpanzees, extinction of the species has now occurred in at least four and seems imminent in five more. All populations are declining in number — world totals may be 150,000 to 250,000. Goodall, on seeing the eroded smallholdings and dwindling fisheries of Lake Tanganyika, laments that "It would be a happy consolation to imagine that such is the price of 'progress', that the human inhabitants of Africa are somehow better off today than they were not so long ago when chimpanzees were a common species — but that is not the case. [Environmental] loss affects us all; human and animal alike."

Africans are responsible for most of the loss of chimpanzees and will be ultimately responsible for saving the wild habitats of these creatures; but Peterson and Goodall choose to play down the mote in the African subsistence farmer's eye and concentrate on the beam in our own. Western nations set the trend for the ever-increasing gobbling of resources — and it is up to us to decide whether chimpanzees should be owned, used and used up. Let it be a reasoned and clear-eyed choice that respects these animals as conscious beings, our sibling species, the demi-humans in our power. □

Alison Jolly is in the Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey, 08544-1003, USA.

Pole polemic

Derek Fordham

The Myth of the Explorer: The Press, Sensationalism and Geographical Discovery. By Beau Riffenburgh. *Belhaven: 1993. Pp. 226. £37.50.*

"A FEW toes aren't much to give to achieve the Pole!" So wrote Robert E. Peary, no doubt with the front page in mind, when several frost-bitten digits fell off in his sock after a disastrous winter journey. This was some years before his last attempt on the North Pole, following which he dispatched the apocryphal cable from Newfoundland: "Stars and Stripes nailed to the Pole". The fact that they were not and that at that moment Frederick Cook was claiming to have reached the Pole a whole year before Peary were issues that whipped the press, which had been supporting Peary, into a state of near apoplexy.

Riffenburgh's scholarly and detailed account of the emergence of the press as a player in the exploration game covers the period 1855–1910, when there were two main arenas in which heroic battles of exploration were fought — Africa and the Arctic. Increasing press interest in exploration ensured that explorers became seen as nationalists fulfilling ideologies, promoted mainly by the press but often, it must be said, by themselves.

This link between press and exploration was forged in the nineteenth century by the James Gordon Bennetts, father and son, of *The New York Herald*. The father claimed to show the world how to obtain news, the son how to create it — a philosophy nowhere more apparent than in the arguments that raged over whether Stanley had actually met Livingstone. This debate became one of the first sources of competition between the American and British press.

Bennett Jr then discovered the Arctic and history might have been quite different had he felt his wallet equal to Stanley's "readiness to undertake the discovery of the North Pole" in 1878. Bennett realized that the public was captivated by reports of struggles to master nature, and he saw that concentrating on the blanks on the map afforded exhilarating possibilities of

establishing man's role as 'Master of Nature'. National heroes were created by means of journalistic hype and thus was established the role of the press in the creation of the modern image of the unknown world, helping to change it from the picturesque vision of the eighteenth century to the sensational perspective that still prevails.

In a manner not unlike that favoured by the press today, the achievements of Amundsen, Nansen and Stanley, among many others, were manipulated and misreported in different but no less effective ways. But of all the issues, African and Arctic, that Riffenburgh examines it is the Arctic and the Cook–Peary con-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

French cartoon of the Peary–Cook flasco in the Arctic, 1909. The caption reads: "Well, if someone's beaten me to the North Pole, I'll have a go at the South. . .".

trovery in particular that is the paradigm of the intrusive role of the press. Such contests as that which arose between Peary and Cook have traditionally been seen as struggles between explorers, but they were in reality equally struggles between their patrons, the newspapers, and were therefore dictated and shaped by the commercial interests of the proprietors.

Peary was a bold, impassioned and eminently dislikeable man who in 1909 set out on what he knew would be his last attempt to reach the North Pole. On this occasion, as in his previous 22 years of polar activity, he was vigorously supported by the press in the form of *The New York Times* and the *National Geographic*. At 87° 47' North, Peary's last supporting party turned back, leaving him as the only capable navigator. Failure now was unthinkable, with

only some 151 nautical miles to the Pole. Peary claimed that daily mileages from this point increased by a factor of four. This outrageous claim has been the focus for the doubts expressed as to whether he reached the Pole (see my reviews in *Nature* **339**, 435 (1989); **344**, 902 (1990) and **348**, 590 (1990)).

Meanwhile, supported by *The New York Herald*, Frederick Cook, Peary's erstwhile colleague, had a year earlier set off on a different route to the Pole. The stage was now set for the controversy that broke when Peary found on his return that Cook claimed to have reached the Pole a year earlier.

This was the sort of issue for which the papers had been waiting. They plunged into the role ascribed by Charles Reich to the modern corporate state: "Utterly out of human control, wholly and perfectly indifferent to any human values". Combat was joined, no holds were barred, nothing was too sensational, and extensive campaigns were mounted to discredit the opposition and increase circulation. Cook was more gentle than Peary and a much more likeable character, 'the Prince of Losers' as he became known. He was unable to produce indisputable evidence of having reached the Pole, but neither could Peary, although with the help of the aggressive campaign against Cook and *The New York Herald* by *The New York Times* and the *National Geographic*, Peary emerged as the victor. Peary understood the press and how it generated images and could deliver fame. He exploited the press as it usually exploited explorers and became established as the hero of an American myth. The general rejection of Cook's claims seriously injured the credibility of *The New York Herald* while Peary's short-lived recognition helped establish *The New York Times* as a journalistic force.

Riffenburgh believes that the image of the explorer has retained a strong hold on the public imagination. The fact that this image has in the main been constructed by the press and often bears little relation to the reality owes more to the public's general preoccupation with tragedy and controversy than with scientific achievement or geographical discovery. The public wanted adventure, not, Stanley claimed, the truth.

The press undoubtedly became expert at manipulating readers' perceptions of explorers. But perhaps explorers since the heyday of the nineteenth century have also been concerned to project a romanticized image of themselves: a recent conqueror of both poles claimed it was all worthwhile, "because it impressed girls at parties!". □

Derek Fordham is at 66 Ashburnham Grove, Greenwich, London SE10 8UJ, UK.

Gallic gloss

Peter Coles

La Science au Présent. Edited by Y. Gautier, D. Laverne, M. Delplanque, F. Gil and J.-L. Verley. *Encyclopaedia Universalis*: 1992. Two-volume set. Pp. 600. FF630.

HERE is a paradox: this two-volume reference should be translated, because it is meant only for French eyes. Aimed at the reader with a good scientific base, it gives a fly-on-the-wall insight into what French scientists worry about. Naturally, this is not what the publisher intended and, I suspect, a French reader would be unaware of how culturally bound the book is. But this is where its real interest lies.

La Science au Présent masquerades as an encyclopaedia of contemporary science and, in a chaotic way, it is. The book covers in detail a range of contemporary issues from high-temperature superconductivity, through cognitive science, to global warming and scientific fraud. Each chapter is written by a specialist, chosen from an army of more than 200 scientists, nearly all from French institutes. But few are well-known outside France, although Axel Kahn, Jean-Marie Lehn and Pierre-Gilles de Gennes make brief appearances.

The quality of writing is uneven. The first chapter of Volume 1, on the search for absolute zero, gets bogged down quickly, whereas much of the second volume has the limpidity of good science journalism. Graphs, photographs and diagrams are put to good use and pithy captions mean one can dip into the book at random, then roam into the text itself for more detail. After all, this is not a book to be read from cover to cover. There is a useful but fairly arbitrary glossary at the end of the second volume, with entries asterisked in the text as they appear. There is also an index to around 1,200 keywords.

So far, nothing surprising. But it soon emerges that *La Science au Présent* is also about French science, with plenty of rallying calls for Gallic researchers 'trampled underfoot by Anglo-Saxon journals' (to paraphrase a statement by the French Science Minister, Hubert Curien, on page 486).

The first signs of idiosyncrasy appear in the way the book is structured. Text-book headings suggest that the gamut of science today can be rendered coherent by the verbs: "observing", "visualizing", "understanding", "questioning", "managing", "producing", "publishing" and "debating". This is the sort of didactic approach exhibition designers use. But the structure is spurious and arbitrary.

And it hampers other kinds of organization, making it difficult just to look something up. To add to the confusion, each author tackles his subject in his own way, so there is no basic grid to fall back on.

Then there is the choice of entries themselves. Of course, there is a good chunk of what preoccupies people at the cutting edge of science worldwide. There is also a large helping of what French scientists — read 'engineers' — are good at: high speed trains, telecommunications, nuclear energy, aerospace, agriculture. A disarming chauvinism creeps in occasionally. Half a page of the glossary is devoted to a list of train speeds, only, one suspects, because France holds the record.

Yet, if the flag-waving is irritating to an English or American reader, this is surely because he or she is unaware of how invasive the equally restricted Anglo-American world-view can be. On close reading, especially of Volume 2, traces of bitterness bubble to the surface. Dominique Stehelin, the Frenchman 'forgotten' from the 1989 Nobel prize for medicine, is reinstated. Montagnier versus Gallo is there (and on page 524 not 526 as N. Witkowski's chapter on fraud says). The glossary contains the esoteric entry "Benveniste (affair)", albeit with a candid description of the debacle that preoccupied *Nature* in late 1988.

The bitterness turns into wishful thinking in what must be a Freudian typographic error. According to a table in Hubert Curien's chapter (page 486), there were exactly as many French biomedical publications (124,096) as American, out of a total of 307,027 in the National Library of Medicine. If this is not a slip, French should be obligatory in all US universities.

But it is chapters like Curien's on the language of science, Martine Barrère's on scientific secrecy, and Geneviève Teil's fascinating analysis of the way in which the style of learned journals promotes an image of the scientist as modest, disinterested and doubting, that should be translated. The general public could be forgiven for thinking the opposite is true today. And — probably no more than paranoia — several French researchers believe American or British reviewers turn down papers because of poor English, then steal the ideas.

Between the lines of this book lie several thought-provoking messages. Flipping through the pictures in the two volumes, one finds *La Science au Présent* to be largely about computers and industrial-scale apparatus. But two-thirds of the world have access to neither. □

Peter Coles is a freelance science writer at 13 rue Titon, 75011 Paris, France.